Applications of Nonstandard Analysis to Markov Processes and Statistical Decision Theory

by

Haosui Duanmu

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Statistical Sciences
University of Toronto

Abstract

Applications of Nonstandard Analysis to Markov Processes
and Statistical Decision Theory

Haosui Duanmu
Doctor of Philosophy
Graduate Department of Statistical Sciences
University of Toronto
2018

We use nonstandard analysis to significantly generalize the well-known Markov chain ergodic theorem and establish a fundamentally new complete class theorem, making progress on two core problems in stochastic process theory and statistical decision theory, respectively.

In the first part, we study the ergodicity of time-homogenous Markov processes. A time-homogeneous Markov process with stationary distribution π is said to be ergodic if its transition probability converges to π in total variation distance. In the most general setting of continuous-time Markov processes with general state spaces, there are few results characterizing the ergodicity of the underlying Markov processes. Using the method of nonstandard analysis, for every standard Markov process $\{X_t\}_{t\geq 0}$, we construct a nonstandard Markov process $\{X'_t\}_{t\in T}$ that inherits most of the key properties of $\{X_t\}_{t\geq 0}$ hence establishing the ergodicity without technical conditions, such as on drift or skeleton chains.

In the second part, we study the relationship between frequentist and Bayesian optimality, extending the line of work initiated by Wald in the 1940's. Existing results are subject to technical conditions that rule out semi-parametric decision problems and generally rule out non-parametric ones. Using nonstandard analysis, we show that, among decision procedures with finite risk functions, a decision procedure is extended admissible if and only if its extension has infinitesimal excess Bayes risk. The result holds in complete generality, i.e, *without* regularity conditions or restrictions on the model or the loss function. This nonstandard characterization of extended admissibility also generates a purely standard theorem: when risk functions are continuous on a compact Hausdorff parameter space, a procedure is extended admissible if and only if it is Bayes.

Acknowledgements

It was a great pleasure to study at University of Toronto. I wish to thank many people who have provided support and help to me.

First of all, I would like to thank my supervisor Jeffrey S. Rosenthal. Jeff provided a relaxing research environment for me to pursue my own interests. He guided me with his extensive knowledge on probability and stochastic processes. My ability as a researcher has improved greatly as a result of his suggestions and insightful comments. Thank you for always having my best interests in mind.

It is difficult to overstate my gratitude to my first co-supervisor William Weiss. With his wealth of knowledge, his inspiration, and his ability to explain abstract mathematical concepts intuitively and precisely, he painted a clear picture of mathematical logic and nonstandard analysis for me. It would have been impossible to finish my dissertation without his dedicated efforts. I adore him both as a mathematician and as a person.

I also owe a great deal to my second co-supervisor Daniel M. Roy for his support and mentorship during my graduate study. Dan provided many brilliant ideas connecting nonstandard analysis and statistics. By collaborating with him, my command of statistics and academic writing improved dramatically. Moreover, Dan provided many chances for me to present my work at conferences and prestigious institutions around the world. It is fair to say that he serves as a role model for me as a researcher.

I wish to thank Professor Michael Evans and Fang Yao for helpful discussions on my research and their thoughtful suggestions for my career. They encouraged me to broaden my views and to discover connections between my research and other related fields.

I must thank my undergraduate NSERC supervisor Franklin D. Tall, with whom I wrote my first research paper. He offered me the opportunity to investigate a long-standing open problem in general topology when I was a third-year undergraduate student. It is with him that I have developed a long-term interest on foundation of mathematics.

The staff in University of Toronto, Department of Statistical Science have provided great support throughout my graduate study. Thank you Andrea Carter, Christine Bulguryemez, Annette Courtemanche, and Angela Fleury for your assistance over the years.

I wish to thank my two best friends from high school (Fei Teng and YuHao Zhong) and my close friends in Toronto (Zhiyan Feng, Peiyang He, Hao Yan, and Yan Yan), for emotional supports, entertainments, and caring they provided. Thank you all for making my life colorful and enjoyable.

Last, and most importantly, I would like to thank my parents, Linling Wang and Qunfan Duanmu, for their unconditional love and consistent support of my interests and ambition. To them I dedicate this thesis.

Contents

Chapter 1

Introduction

During the period of 1957–1965, Abraham Robinson introduced nonstandard analysis, a formal framework built on mathematical logic in which one can rigorously define infinitesimal and infinite numbers. Nonstandard analysis has advanced rapidly since its introduction by Robinson, with much of this progress driven by applications to new areas of mathematics, especially probability theory. However, due to the use of mathematical logic, the proportion of mathematicians who use nonstandard analysis effectively in research is, and always has been, infinitesimal. As a result, the potential impact of nonstandard analysis has not been fully realized. In this dissertation, we will illustrate the power of nonstandard analysis by significantly generalizing the well-known Markov chain ergodic theorem and establishing a fundamentally new complete class theorem, making progress on two core problems in stochastic process theory and statistical decision theory, respectively.

Nonstandard models are constructed to satisfy the following three principles:

1. *extension*, associating every standard mathematical object with a nonstandard mathematical object called its extension;

2. *transfer*, allowing us to use first-order logic to make connections between standard and nonstandard object; and

3. *saturation*, giving us a powerful mechanism for proving the existence of nonstandard objects.

The formal definitions of these three principles are easily understood but the consequences are far reaching. Indeed, all the results in this dissertation involving nonstandard analysis are consequences of effective applications of these three principles.

The power of nonstandard analysis comes from its ability to link finite/discrete with the infinite/continuous.

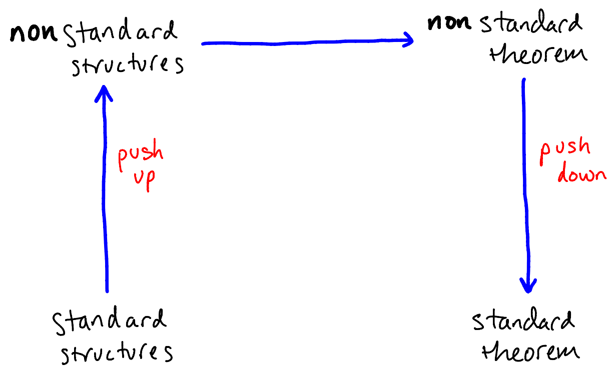

Figure 1.1: The structure of a "push up/down" argument in nonstandard analysis. (Image courtesy of Daniel Roy.)

One way to establish such a link is via *hyperfinite objects*. Roughly speaking, hyperfinite objects are infinite objects which possesses all the first-order logic properties of finite objects. Hyperfinite objects can be used to represent standard infinite mathematical objects. For example, Henson [17] and Anderson [2] show that, under moderate assumptions, every probability measure can be "represented" by a nonstandard probability measure with hyperfinite support. As a concrete example, Lebesgue measure λ on $[0,1]$ can be replaced in many situations by the uniform distribution on $\{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}, 1\}$, where N is an infinitely large natural number. As for a more sophisticated example, Anderson [1] showed that Brownian motion can be represented by a hyperfinite random walk with an infinitesimal increment.

In the other direction, we can often construct standard mathematical objects from hyperfinite ones. Thus, nonstandard analysis provides a general methodology to solve standard mathematical problems. The general structure of this approach is the following (see Fig. 1.1): Consider an existing mathematical theorem involving one or more finite objects. In order to establish an analogous result for infinite objects, we can search for hyperfinite approximations of these infinite objects and use the properties of hyperfinite sets to establish a hyperfinite counterpart of the original theorem. Under regularity conditions, we may then be able to "push down" the hyperfinite result to obtain a standard theorem. Thus, this general approach can be used to solve mathematical problems involving infinite objects provided that the finite case is well-understood.

1.1 Applications to Probability Theory and Statistics

In this dissertation, we study Markov chain ergodic theorems in probability theory and complete class theorems in statistical decision theory. In both theories, finite/discrete theorems are well-understood.

Using nonstandard analysis, we establish hyperfinite counterparts of both theorems. Neither is a trivial application of the transfer principle—saturation is essential. We then apply push-down techniques to establish infinite/continuous versions of a Markov chain ergodic theorem and a complete class theorem. Both theorems are new results.

1.1.1 Markov Chain Ergodic Theorem

A Markov process is ergodic if its transition probability converges to its stationary distribution in total variation distance. The ergodicity of Markov processes is of fundamental importance in the study of Markov processes. On one hand, the ergodicity of a Markov process allows us to disregard the initial distribution of the Markov process and replace its n-step transition probability by the stationary distribution for n large enough. On the other hand, in the Markov chain Monte Carlo context, one can sample form the n-step transition distribution instead of sampling from the stationary distribution for large n. The Markov chain ergodic theorem is well-known for Markov processes with discrete time-line and countable state space (see e.g., [7, 15, 42]). However, for processes in continuous time and space, there is no such clean result; the closest are apparently the results in [31–33] using complicated assumptions about skeleton chains together with drift conditions (see Theorem 5.3.7). Other existing results (see e.g., [51]) make extensive use of the techniques are results from [32, 33].

Meanwhile, nonstandard analysis provides an alternative way to study general stochastic processes by associating every standard stochastic process with a hyperfinite stochastic process. Anderson [1] gave a nonstandard construction of the Brownian motion and the Itô integral. In particular, he showed that the Brownian motion can be represented as a hyperfinite random walk with infinitesimal increment. Keisler [23] used Anderson's result as the starting point for a deep study of stochastic differential equations and Markov processes. In this dissertation, we generalize Anderson's work to give a hyperfinite representation for continuous-time general state space Markov processes satisfying certain regularity conditions. We also give a proof of the Markov chain ergodic theorem in a very general setting.

Given a continuous-time general state space Markov process $\{X_t\}_{t\geq 0}$, under moderate regularity conditions, we associate it with a hyperfinite Markov process $\{X'_t\}_{t\in T}$, that is, a Markov process with hyperfinite state space and hyperfinite time-line. To construct $\{X'_t\}_{t\in T}$, we first define the time-line T to be $\{0,\delta t,2\delta t,\ldots,K\}$ for some positive infinitesimal δt and some positive infinite number K. We then partition the nonstandard extension of the state space of $\{X_t\}_{t\geq 0}$ into hyperfinitely many pieces of nonstandard Borel sets with infinitesimal radius and pick one "representative" point from each piece to form the hyperfinite state space $S=\{s_1,s_2,\ldots,s_N\}$ of $\{X'_t\}_{t\in T}$. For $s_i,s_j\in S$, the one-step transition probability from s_i to s_j is defined to be the nonstandard transition probability from s_i to $B(s_j)$ at time δt, where $B(s_j)$ denotes the nonstandard Borel set containing s_j. It can be shown that the nonstandard transition probability of $\{X'_t\}_{t\in T}$ differs from the transition probability of $\{X_t\}_{t\geq 0}$ by only infinitesimal, hence $\{X'_t\}_{t\in T}$ provides a robust approximation of $\{X_t\}_{t\geq 0}$

Meanwhile, due to the similarity between hyperfinite objects and finite objects, $\{X'_t\}_{t\in T}$ satisfies the same first-order logic properties as Markov processes with discrete time-line and finite state space. Thus, we can establish the ergodicity of $\{X'_t\}_{t\in T}$ by mimicking the proof of the Markov chain ergodic theorem for discrete-time Markov processes with finite state spaces. Finally, we show that, under moderate regularity conditions, the ergodicity of $\{X'_t\}_{t\in T}$ implies the ergodicity of $\{X_t\}_{t\geq 0}$, establishing the Markov chain ergodic theorem for continuous-time general state space Markov processes.

1.1.2 Statistical Decision Theory

Statistical decision theory provides a formal framework in which to study the process of making decisions under uncertainty. Statistical decision theory was introduced in 1939 by Wald, who noted that many hypotheses testing and parameter estimation could be considered as special cases of his general notion of decision problems. Since its introduction, statistical decision theory has served as a rigorous foundation of statistics for over half of a century. In this dissertation, we are interested in studying the deep connection between frequentist notions (in particular, admissibility and extended admissibility) and Bayesian optimality.

A decision procedure is inadmissible if there exists another procedure whose risk is everywhere no worse and somewhere strictly better. Ignoring issues of computational complexity, one should never use an inadmissible decision procedure. Thus, admissibility is necessary condition for any reasonable notion of optimality.

It has long been known that there are deep connections between admissibility and Bayes optimality. In one direction, under suitable regularity conditions, every admissible procedure is Bayes with respect to a carefully chosen prior, improper prior, or sequence thereof. The resulting (quasi-)Bayesian interpretation provides insight into the strengths and weaknesses of the procedure from an average-case perspective. In the other direction, (necessary and) sufficient conditions for admissibility expressed in terms of (generalized) priors point us towards Bayesian procedures with good frequentist properties.

For statistical decision problems with finite parameter spaces, it is well-known that a decision procedure is extended admissible if and only if it is Bayes (see e.g., [14, 26]). For statistical decision problems with infinite parameter spaces, on the other hand, there exists an admissible decision procedure which is not Bayes. Thus, one must relax the notion of Bayesian optimality to regain a tight link between frequentist and Bayesian optimality (see e.g., [5, 9, 10, 20, 25, 43, 50, 52, 54–57]). As the literature stands, for statistical decision problems with infinite parameter spaces, connections between frequentist and Bayesian optimality are subject to regularity conditions, and these conditions often rule out semi-parametric and nonparametric problems. As a result, the relationship between frequentist and Bayesian optimality in the setting of modern statistical decision problems is often uncharacterized.

In contrast to existing methods in the literature, nonstandard analysis offers a different approach in solving this long-standing open problem. Informally speaking, the utility of nonstandard models for statistical decision theory stems from two sources: first, every nonstandard model possesses nonstandard reals numbers, including infinitesimal / infinite positive numbers which can be used to construct priors to make extreme statement, e.g., priors assigning positive but infinitesimal mass to some points. Using these priors, we are able to form a nonstandard version of Bayesian optimality and are able to establish the equivalence between frequentist and Bayesian optimality without any regularity conditions.

In particular, using a separating hyperplane argument in concert with the three principles outline in nonstandard analysis (extension, transfer and saturation), we show that a standard decision procedure δ is extended admissible if and only if, for some nonstandard prior, the Bayes risk of its extension ${}^*\delta$ is within an infinitesimal of the minimum Bayes risk among all extensions. Such a decision procedure is said to be nonstandard Bayes. For any metric on the parameter space Θ such that risk functions are continuous, we are able to show that a procedure is admissible if its extension is nonstandard Bayes with respect to a prior that assigns sufficient mass to every standard open ball. The result is a nonstandard variant of Blyth's method, in which a sequence of priors is replaced by a single nonstandard prior in order to witness admissibility. We also apply our nonstandard theory to give a purely standard result: On compact Hausdorff parameter spaces when risk functions are continuous, a decision procedure is extended admissible if and only if is Bayes.

1.2 Overview of the Dissertation

We conclude with a chapter-by-chapter summary: In Chapter 1, we develop, from the beginning, the notions needed from nonstandard analysis, including the three basic principles, the standard part map, internal sets, hyperfinite sets, and Loeb measures. We then discuss various sufficient conditions under which the standard part map is measurable. We close with a general discussion on hyperfinite representations of standard probability spaces.

We start Chapter 2 by introducing hyperfinite Markov processes, and then prove a hyperfinite Markov chain ergodic theorem in Section 3.1. In Sections 3.2 and 3.3, we give explicit constructions of hyperfinite representations for discrete-time general state space Markov processes and continuous-time general state space Markov processes, respectively.

In Chapter 3, under moderate regularity conditions, we establish the Markov chain ergodic theorem for continuous-time Markov processes with general state spaces using results from Chapter 2. For a continuous-time general state space Markov process $\{X_t\}_{t\geq 0}$, we first establish the ergodicity of its hyperfinite representation $\{X'_t\}_{t\in T}$ then apply "push-down" techniques to establish ergodicity of $\{X_t\}_{t\in T}$.

In Chapter 4, we discuss constructions of standard Markov processes and stationary distributions from hyperfinite Markov processes. We close with remarks and open problems related to Markov chains.

In Chapter 5, we begin our study of statistical decision theory by introducing its basic concepts and discussing connections between admissibility, Bayes optimality, and complete classes. We close with an extensive literature review of existing results on complete classes.

In Chapter 6, we study the nonstandard extensions of decision problems and define a novel notion of nonstandard Bayes optimality. We then show that a decision procedure is extended admissible if and only if its nonstandard extension is nonstandard Bayes, i.e., its Bayes risk is within an infinitesimal of the minimum Bayes risk among all extensions. This result holds in complete generality.

In Chapter 7, we give sufficient nonstandard conditions for admissibility of a standard decision procedure. We also establish a standard result: For decision problems with compact parameter space and continuous risk functions, a decision procedure is extended admissible if and only if it is Bayes. Finally we close with remarks and open problems in statistical decision theory.

We will assume that the reader is familiar with measure-theoretic probability theory, and has had some basic exposure to statistics and mathematical logic. For background material on nonstandard analysis, see [40], [3], [11], and [58]. For background on Markov processes, see [42] and [31]. For background on statistical decision theory, see [14] and [26].

Chapter 2

Nonstandard Analysis and Internal Probability Theory

This dissertation uses Robinson's nonstandard analysis to study fundamental problems in statistics and probability theory. Nonstandard analysis is introduced by Abraham Robinson in [40]. A comprehensive account of modern nonstandard analysis is contained in [3] and [11]. In this chapter, we develop from the beginning the knowledge and notions needed from nonstandard analysis.

We start by introducing some basic notions in nonstandard analysis, including superstructures, internal and external sets, the transfer and the saturation principle. For construction of the nonstandard universe, interested readers can read [3, Section. 1]. In Section 2.1.1, we investigate basic properties of the nonstandard real line, ${}^*\mathbb{R}$, which is undoubtedly the most well-known nonstandard object. We extend most of the notions and properties on ${}^*\mathbb{R}$ to general topological (metric) spaces in Section 2.1.2.

In Section 2.2, we give an introduction to nonstandard measure theory. The nonstandard measure theory is formulated by Peter Loeb in his landmark paper [28]. In [28], Loeb constructed a standard countably additive probability space (called the Loeb space) which is the completion of a nonstandard probability space (called an internal probability space). We start Section 2.2 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. A hyperfinite set is an infinite set 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 "good representations" for standard probability spaces. We illustrate this idea in Example 2.2.5 and the remark after it. We also discuss nonstandard product measure and nonstandard integration theory in this section.

In Section 2.3, we discuss the measurability of the standard part map. A nonstandard element x is *near-standard* if there is a standard element x_0 which is infinitely close to it. Such x_0 is called the *standard part* of x. The standard part map st maps a near-standard nonstandard element to a its standard part. The connection between a standard probability space and its nonstandard extension (which is an internal probability space) can usually be established via studying the standard part map. Thus, it is natural to require that st 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 [24], it has been shown that the answer of this question largely depend on the Loeb measurability of $\mathrm{NS}({}^*X) = \{x \in {}^*X : (\exists y \in X)(y = \mathsf{st}(x))\}$ (the collection of all near-standard points in ${}^*X$). In [3, 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 2.3.5. We are also able to obtain a stronger result by assuming the space is merely Cech-complete (see Theorem 2.3.6).

In Section 2.4, we discuss the idea of using hyperfinite probability spaces to represent standard probability spaces. Such hyperfinite probability space is called a hyperfinite representation of the underlying standard probability space. We restrict our attention to σ-compact metric spaces satisfying the Heine-Borel condition. In Definition 2.4.3, we give the definition of hyperfinite representations of a σ-compact metric space X satisfying the Heine-Borel condition. The idea is to decompose X into hyperfinitely many *Borel sets with infinitesimal diameters and pick one point 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 for $\{B(s) : s \in S\}$ to cover ${}^*X$. Thus, we only require $\{B(s) : s \in S\}$ to cover a "large enough" portion of ${}^*X$. A hyperfinite representation S has two parameters r and ε. The parameter r measures the portion of ${}^*X$ that is covered by $\{B(s) : s \in S\}$ while ε puts an upper bound on the diameters of the elements in $\{B(s) : s \in S\}$. Given an (ε, r)-hyperfinite representation S, in Theorem 2.4.11, we define an internal probability measure P' on $(S, \mathscr{I}[S])$ and establishes the link between $(X, \mathscr{B}[X], P)$ and $(S, \mathscr{I}(S), P')$. Theorem 2.4.11 is similar to [11, Theorem 3.5 page 159] which was proved in [2].

2.1 Basic Concepts in Nonstandard Analysis

Those familiar with nonstandard methods may safely skip this section on their first reading. Nonstandard analysis is introduced by Abraham Robinson in [40]. For modern applications of nonstandard analysis, interested readers can read [3] or [11]. Our following introduction of nonstandard analysis owes much to [3].

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 $\mathscr{L}(\mathbb{V}(S))$ 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 $\mathscr{L}(\mathbb{V}(S))$ are defined recursively:

- If x and y are variables and a and b are constants,

  $(x = y), (x \in y), (a = x), (a \in x), (x \in a), (a = b), (a \in b)$ 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 $\mathscr{L}(\mathbb{V}(S))$: 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 symbol 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 2.1.1. 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 $\mathscr{L}(\mathbb{V}(S))$, 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 $\mathscr{F} = \{A_i : i \in I\}$ of internal sets indexed by a set I of cardinality less than κ, if $\mathscr{F}$ has the finite intersection property, i.e., if every finite intersection of elements in $\mathscr{F}$ is nonempty, then the total intersection of $\mathscr{F}$ is non-empty.

A $\aleph_1$ saturated model can be constructed via an ultrafilter, see [3, Thm. 1.7.13].

The language of $\mathbb{V}({}^*S)$ is almost the same as $\mathscr{L}$ 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 $\mathscr{L}(\mathbb{V}({}^*S))$. If $\phi(x_1, \dots, x_n)$ is a formula in $\mathscr{L}(\mathbb{V}(S))$ with free variables $x_1, \dots, x_n$, then the $*$-transfer of ϕ is the formula in $\mathscr{L}(\mathbb{V}({}^*S))$ 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 2.1.2. 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 2.1.3 ([29])**.** *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 2.1.2, 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 2.1.4. *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 2.1.5. *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 2.1.6 (Internal Definition Principle)**.** *Let $\phi(x)$ be a formula in $\mathscr{L}(\mathbb{V}({}^*S))$ 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 2.1.7. 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 2.1.8 ([3, 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 2.1.9. 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 2.1.1.

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 2.1.10. 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 2.1.11 ([3])**.** *A subset A of a hyperfinite set T is internal if and only if A is hyperfinite.*

An immediate consequence of Theorem 2.1.8 is:

Proposition 2.1.12 ([3, 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.*

2.1.1 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 2.1.13. 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 2.1.14. 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 2.1.15. *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 2.1.16. 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 2.1.17. 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 2.1.18. 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 2.1.19. *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 2.1.20. *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 $(a_1b_1 + a_2b_2 + \cdots + a_Nb_N) - (a_1c_1 + a_2c_2 + \cdots + a_Nc_N) = a_1(b_1 - c_1) + a_2(b_2 - c_2) + \cdots + a_N(b_N - c_N) \leq \max\{a_i|b_i - c_i| : i \leq N\} \approx 0$. □

2.1.2 Nonstandard Extensions of General Metric Spaces

We generalize the concepts developed in Section 2.1.1 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 $\mathscr{B}_x$ denote a local base at point x.

Definition 2.1.21. Given $x \in X$, the monad of x is

$$\mu(x) = \bigcap_{U \in \mathscr{B}_x} {}^*U. \tag{2.1.1}$$

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 2.1.22. 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 2.1.23. 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 2.1.24. 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 2.1.25 ([3, 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 $\mathscr{U} = \{O_\alpha : \alpha \in \mathscr{A}\}$ be an open cover of A with no finite subcover. By Proposition 2.1.12, let $\mathscr{B}$ be a hyperfinite collection of ${}^*\mathscr{U}$ containing ${}^*O_\alpha$ for all $\alpha \in \mathscr{A}$.

By the transfer principle, there exists a $y \in {}^*A$ such that $y \notin U$ for all $U \in \mathscr{B}$. Thus, $y \notin {}^*O_\alpha$ for all $\alpha \in \mathscr{A}$. 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 2.1.26. Consider ${}^*[0,1] = \{x \in {}^*\mathbb{R} : 0 \le x \le 1\}$, 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 $\varepsilon \in {}^*\mathbb{R}$. Then $\varepsilon \in {}^*(0,1)$ but $\varepsilon \notin \mathrm{NS}({}^*(0,1))$.

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

Theorem 2.1.27 ([29])**.** *Let $(X, \mathscr{T})$ be a regular Hausdorff space. Suppose the nonstandard extension is more saturated than the cardinality of $\mathscr{T}$. 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 2.1.25, 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. □

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

Theorem 2.1.28. *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 $K_1 = \bigcup_{i \le n} C_{x_i}$. 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 2.1.29. 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 ${}^*K_n$.

2.2 Internal Probability Theory

In this section, we give a brief introduction to nonstandard probability theory. The interested reader can consult [23] and [3, Section 4] for more details. The expert may safely skip this section on first reading.

Let Ω be an internal set. An *internal algebra* $\mathscr{A} \subset \mathscr{P}(()\Omega)$ is an internal set containing Ω and closed under taking complement and hyperfinite unions/intersections. A set function $P \colon \mathscr{A} \to {}^*\mathbb{R}$ is *hyperfinitely additive* when, for every $n \in {}^*\mathbb{N}$ and mutually disjoint family $A_1, \dots, A_n \in \mathscr{A}$, 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 2.2.1. An internal finitely-additive probability space is a triple $\{\Omega, \mathscr{A}, P\}$ where:

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

Example 2.2.2. Let $(X, \mathscr{A}, P)$ be a standard probability space. Then $({}^*X, {}^*\mathscr{A}, {}^*P)$ is an internal probability space. Although $\mathscr{A}$ is a σ-algebra and P is countably additive, $\mathscr{A}$ 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 2.2.3. A hyperfinite probability space is an internal probability space $(\Omega, \mathscr{A}, P)$ where:

1. Ω is a hyperfinite set.

2. $\mathscr{A} = \mathscr{I}(\Omega)$ where $\mathscr{I}(\Omega)$ 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 [28] 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 2.2.4 ([28])**.** *Let $(\Omega, \mathscr{A}, P)$ be an internal finitely additive probability space; then there is a standard (σ-additive) probability space $(\Omega, \overline{\mathscr{A}}, \overline{P})$ such that:*

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

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

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

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

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

1. $\overline{\mathscr{A}}$ equals to:

$$\{A \subset \Omega | \forall \varepsilon \in \mathbb{R}^+ \exists A_i, A_o \in \mathscr{A} \text{ such that } A_i \subset A \subset A_o \text{ and } P(A_o \setminus A_i) < \varepsilon\}. \tag{2.2.1}$$

2. For all $A \in \overline{\mathscr{A}}$ we have:

$$\overline{P}(A) = \inf\{\overline{P}(A_o) | A \subset A_o, A_o \in \mathscr{A}\} = \sup\{\overline{P}(A_i) | A_i \subset A, A_i \in \mathscr{A}\}. \tag{2.2.2}$$

In fact, the Loeb σ-algebra can be taken to be the $\overline{P}$-completion of the smallest σ-algebra generated by $\mathscr{A}$. 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 2.2.5. Let $(\Omega, \mathscr{A}, P)$ 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 $\mathscr{A} = \mathscr{I}(\Omega)$(Recall that $\mathscr{I}(\Omega)$ is the collection of all internal subsets of Ω). Define P on $\mathscr{A}$ by letting $P(\omega) = \delta t$ for all $\omega \in \Omega$. This is called the uniform hyperfinite Loeb measure.

Claim 2.2.6. $\mathsf{st}^{-1}(0) \cap \Omega \in \overline{\mathscr{A}}$

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 $A_n \in \mathscr{A}$ for all $n \in \mathbb{N}$. Hence $\bigcap_{n\in\mathbb{N}} A_n \in \overline{\mathscr{A}}$. Thus $\mathsf{st}^{-1}(0) \cap \Omega = \bigcap_{n\in\mathbb{N}} A_n \in \overline{\mathscr{A}}$, completing the proof. □

Let ν denote the Lebesgue measure on $[0,1]$. In Section 2.4, 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 $(\Omega, \mathscr{A}, P)$ to represent the Lebesgue measure on $[0,1]$. $(\Omega, \mathscr{A}, P)$ 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 2.4.

As $\mathsf{st}^{-1}(0)$ is an external set, Example 2.2.5 shows that the Loeb σ-algebra contains external sets.

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

Definition 2.2.7. The product Loeb measure $\overline{P}_1 \times \overline{P}_2$ is defined to be the probability measure on $(\Omega \times \Gamma, \overline{\mathscr{A}} \otimes \overline{\mathscr{D}})$ satisfying:

$$(\overline{P}_1 \times \overline{P}_2)(A \times B) = \overline{P}_1(A) \cdot \overline{P}_2(B). \tag{2.2.3}$$

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

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

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

Definition 2.2.8. The product internal measure $P_1\times P_2$ is defined to be the internal probability measure on $(\Omega\times\Gamma,\mathscr{A}\otimes\mathscr{D})$ satisfying:

$$(P_1\times P_2)(A\times B)=P_1(A)\cdot P_2(B). \tag{2.2.4}$$

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

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

Example 2.2.9. Suppose both $(\Omega,\mathscr{A},P_1)$ and $(\Gamma,\mathscr{D},P_2)$ are hyperfinite probability spaces. Recall from Definition 2.2.3 that $\mathscr{A}=\mathscr{I}(\Omega)$ and $\mathscr{D}=\mathscr{I}(\Gamma)$ where $\mathscr{I}(\Omega)$ and $\mathscr{I}(\Gamma)$ 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 $\mathscr{I}(\Omega\times\Gamma)$. 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 $(\Omega\times\Gamma,\mathscr{A}\otimes\mathscr{D},P_1\times P_2)$, the Loeb construction can be applied to give a Loeb probability space $(\Omega\times\Gamma,\overline{(\mathscr{A}\otimes\mathscr{D})},\overline{(P_1\times P_2)})$. It is natural to seek for relation between $(\Omega\times\Gamma,\overline{(\mathscr{A}\otimes\mathscr{D})},\overline{(P_1\times P_2)})$ and $(\Omega\times\Gamma,\overline{\mathscr{A}}\otimes\overline{\mathscr{D}},\overline{P}_1\times\overline{P}_2)$.

Theorem 2.2.10 ([23])**.** *Consider two Loeb probability spaces $(\Omega,\overline{\mathscr{A}},\overline{P}_1)$ and $(\Gamma,\overline{\mathscr{D}},\overline{P}_2)$. We have $\overline{(P_1\times P_2)}=\overline{P}_1\times\overline{P}_2$ on $\overline{\mathscr{A}}\otimes\overline{\mathscr{D}}$.*

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

$$(P_1\times P_2)((A_o\times B_o)\setminus(A_i\times B_i))=(P_1\times P_2)((A_o\setminus A_o)\times(A_i\setminus B_i))=\varepsilon^2<\varepsilon. \tag{2.2.5}$$

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

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

We then have:

$$\overline{P}_1 \times \overline{P}_2(A \times B) \tag{2.2.6}$$

$$= \sup\{\mathrm{st}(P_1(A_i))|A_i \subset A, A_i \in \mathscr{A}\} \times \sup\{\mathrm{st}(P_2(B_i))|B_i \subset A, B_i \in \mathscr{D}\} \tag{2.2.7}$$

$$= \sup\{\mathrm{st}(P_1(A_i))\mathrm{st}(P_2(B_i))|A_i \subset A, A_i \in \mathscr{A}, B_i \subset A, B_i \in \mathscr{D}\} \tag{2.2.8}$$

$$= \overline{(P_1 \times P_2)}(A \times B), \tag{2.2.9}$$

completing the proof. □

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

Example 2.2.11. [23] Let Ω be an infinite hyperfinite set. Let $\Gamma = \mathscr{I}(\Omega)$. Let $(\Omega, \mathscr{I}(\Omega), P)$ and $(\Gamma, \mathscr{I}(\Gamma), Q)$ be two uniform hyperfinite probability spaces over the respective sets. Let $E = \{(\omega, \lambda) : \omega \in \lambda \in \Gamma\}$. It can be shown that $E \in \overline{(\mathscr{I}(\Omega) \otimes \mathscr{I}(\Gamma))}$ but $E \notin \overline{\mathscr{I}(\Omega)} \otimes \overline{\mathscr{I}(\Gamma)}$.

In fact, it can be shown that $(\overline{P \times Q})(E) > 0$ while $\overline{P}(A)\overline{Q}(B) = 0$ for every $A \in \mathscr{I}[\Omega]$ and every $B \in \mathscr{I}[\Gamma]$. However, the internal probability space $(\Gamma, \mathscr{I}[\Gamma], Q)$ does not corresponds to any standard probability space.

Open Problem 1. *Let $(\Omega, \mathscr{A}, P)$ be an internal probability space. Let $(\overline{P \times P})(B) > 0$ for some $B \in \overline{\mathscr{A} \otimes \mathscr{A}}$. Under what conditions does there exists $C \in \overline{\mathscr{A}} \otimes \overline{\mathscr{A}}$ such that $C \subset B$ and $(\overline{P \times P})(C) > 0$? Does $(\Omega, \mathscr{A}, P)$ being the nonstandard extension of some standard probability space help?*

2.2.2 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 2.2.12 ([3, 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 $\mathrm{st}(F)$ exists everywhere. Then $\mathrm{st}(F)$ 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 2.2.13 ([3, 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 2.2.14. 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 2.2.13, 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 2.2.15. 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 2.2.16 ([3, 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.*

2.3 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 $E \in \mathscr{B}[X]$ 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,. [3, 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 [3, 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 [18] is a key result regarding the measurability of st.

Theorem 2.3.1 ([3, 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 $({}^*X, {}^*\mathscr{B}[X])$ and suppose $\mathrm{NS}({}^*X) \in \overline{{}^*\mathscr{B}[X]}$; then st is Borel measurable from $({}^*X, \overline{{}^*\mathscr{B}[X]})$ to $(X, \mathscr{B}[X])$.*

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

We first recall the following definitions from general topology.

Definition 2.3.2. 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 2.3.3. 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 2.3.4 ([24])**.** *Suppose that $(\Omega, \mathscr{A}, P)$ is an internal finitely additive probability space with corresponding Loeb space $(\Omega, \mathscr{A}_L, \overline{P})$ and suppose that $\mathscr{C}$ is a subset of $\mathscr{A}$ such that the nonstandard model is more saturated than the external cardinality of $\mathscr{C}$. Then $\bigcap \mathscr{C} \in \mathscr{A}_L$. Furthermore, if $P(A) = 1$ for all $A \in \mathscr{C}$, then $\overline{P}(\bigcap \mathscr{C}) = 1$*

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

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

In the context of Lemma 2.3.4, by considering the complement, it is easy to see that $\bigcup \mathscr{C} \in \overline{\mathscr{A}}$. Similarly, if we have $P(A) = 0$ for all $A \in \mathscr{C}$ then $\overline{P}(\bigcup \mathscr{C}) = 0$.

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

Lemma 2.3.5 ([24]). *Let X be a σ-compact space with Borel σ-algebra $\mathscr{B}[X]$ and let $({}^*X, {}^*\mathscr{B}[X]_L, \overline{P})$ be a Loeb space. Then $\mathrm{NS}({}^*X) \in \overline{{}^*\mathscr{B}[X]}$.*

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

Theorem 2.3.6. *If the Tychnoff space X is Cech complete then $\mathrm{NS}({}^*X) \in {}^*\mathscr{B}[X]_L$.*

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

$$ {}^*Y = {}^*X \cup {}^*S = \mathrm{NS}({}^*X) \cup {}^*S \cup Z. \tag{2.3.1} $$

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 2.3.7. *For any $i \in \omega$, $N_i \in \overline{{}^*\mathscr{B}[X]}$.*

Proof. : Without loss of generality, it is enough to prove the claim for N_1. Let $\mathscr{U} = \{U \subset X$: U is open and $S_1 \subset U\}$. We claim that $N_1 = \bigcap\{{}^*U : U \in \mathscr{U}\}$. To see this, we first consider any $u \in \bigcap\{{}^*U : U \in \mathscr{U}\}$. 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 $\mathscr{U}$. Hence we have a contradiction. Conversely, it is easy to see that $N_1 \subset \bigcap\{{}^*U : U \in \mathscr{U}\}$. We also know that each ${}^*U \in {}^*\mathscr{B}[X]$. Assume that we are working on a nonstandard extension which is more saturated than the cardinality of the topology of X, then for any $i \in \omega$ $N_i \in \overline{{}^*\mathscr{B}[X]}$ by Lemma 2.3.4. $\square$

It is also easy to see that $\bigcup_{i<\omega} N_i = \mathrm{NS}({}^*S) \cup Z$. By Lemma 2.3.5, we know that both $\bigcup_{i<\omega} N_i$ and $\mathrm{NS}({}^*S)$ belong to $\overline{{}^*\mathscr{B}[Y]}$. Hence $Z \in \overline{{}^*\mathscr{B}[Y]}$.

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

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

Thus, by Theorem 2.3.1, 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 2.2 that given an internal algebra $\mathscr{A}$ its Loeb extension $\overline{\mathscr{A}}$ is actually the $\overline{P}$-completion of the σ-algebra generated by $\mathscr{A}$. So $\mathscr{A}_L$ could differ for different internal probability measures. We use $\overline{\mathscr{A}}^P$ to denote the Loeb extension of $\mathscr{A}$ with respect to the internal probability measure P.

Definition 2.3.8. A set $A \subset {}^*X$ is called universally Loeb-measurable if $A \in \overline{\mathscr{A}}^P$ for every internal probability measure P on $({}^*X, \mathscr{A})$.

We denote the collection of all universally-Loeb measurable sets by $\mathscr{L}(\mathscr{A})$. By Theorem 2.3.6, $\mathrm{NS}({}^*X)$ is universally Loeb measurable if X is Cech complete. Moreover, Theorem 2.3.1 can be restated as following:

Theorem 2.3.9 ([24])**.** *Let X be a Hausdorff regular space equipped with Borel σ-algebra $\mathscr{B}[X]$. If $B \in \mathscr{B}[X]$ then $\mathsf{st}^{-1}(B) \in \{A \cap \mathrm{NS}({}^*X) : A \in \mathscr{L}(\mathscr{B}[X])\}$.*

Thus, by Theorem 2.3.6, $\mathsf{st}^{-1}(B)$ is universally measurable for every $B \in \mathscr{B}[X]$ 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 2.3.10. *[3, Example 4.1] There is a separable metric space X and a Loeb space $({}^*X, {}^*\mathscr{B}[X], \overline{P})$ 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 $([0,1], \mathscr{B}[[0,1]])$. Let $\mathscr{A}$ be the σ-algebra generated from $\mathscr{B}[[0,1]] \cup \{X\}$. Let m be the extension of μ to $\mathscr{A}$ by letting $m(X) = 1$.

Claim 2.3.11. *m is a probability measure on $([0,1], \mathscr{A})$.*

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

$$m(A \cap X) = m(B \cap X) \to m(A) = m(B). \tag{2.3.2}$$

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 ${}^*\mathscr{B}[X]$. Consider the internal probability space $({}^*X, {}^*\mathscr{B}[X], P)$. Let $A \in \mathrm{NS}({}^*X) \cap {}^*\mathscr{B}[X]$ and let $A' = \mathsf{st}_X(A)$ where $\mathsf{st}_X(A) = \{x \in X : (\exists a \in A)(a \approx x)\}$. By Theorem 2.1.27, 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 $\varepsilon > 0$, there exists an open set $U_\varepsilon \subset [0,1]$ of Lebesgue measure less than ε that contains A'. As $A' = \mathsf{st}_X(A)$, we know that $A \subset {}^*X \cap {}^*U_\varepsilon \subset {}^*U_\varepsilon$. Then $P(A) \leq {}^*m({}^*U_\varepsilon) < \varepsilon$. 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 [3, Example 4.1]. In [3, 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.

2.4 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 [2, 6, 17, 27]. In this section, we establish a hyperfinite representation theorem for σ-compact complete 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 2.4.1. 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 2.1.2, finite elements of complete metric spaces need not be near-standard. However, finite elements are near-standard for σ-compact metric spaces satisfying the Heine-Borel condition.

Theorem 2.4.2. *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 2.1.25, 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 2.1.25, 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. □

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 2.4.3. Let X be a σ-compact complete metric space satisfying the Heine-Borel condition. Let $\varepsilon \in {}^*\mathbb{R}^+$ 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 $B(s) \in {}^*\mathscr{B}[X]$ 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

$$\mathrm{NS}({}^*X) \subset \bigcup_{s \in S} B(s) = \overline{U(a_0, r_0)} \tag{2.4.1}$$

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

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 2.4.3 is *compact.

Example 2.4.5. Consider the real line $\mathbb{R}$ with standard metric. Fix $N_1, N_2 \in {}^*\mathbb{N} \setminus \mathbb{N}$. Let $\varepsilon = \frac{1}{N_1}$ and let $r = 2N_2$. It then follows that

$$S = \{-2N_2, -2N_2 + \frac{1}{N_1}, \dots, -\frac{1}{N_1}, 0, \frac{1}{N_1}, \dots, 2N_2\} \tag{2.4.2}$$

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

To see this, we need to check the three conditions in Definition 2.4.3. 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 2.4.6. *Let X be a σ-compact complete 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 $\varepsilon_0 \in {}^*\mathbb{R}^+$ and an infinite r_0. Pick any standard $x_0 \in X$ and consider the open ball

$$U(x_0, 2r_0) = \{x \in {}^*X : {}^*d(x, x_0) < 2r_0\}. \tag{2.4.3}$$

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:

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

By the transfer principle, we have:

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

Let $\mathscr{A} = \{U_i : i \leq K\}$. 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 $S_{\varepsilon_0}^{r_0} = \{s_i : i \leq K\}$ and it is easy to check that this $S_{\varepsilon_0}^{r_0}$ satisfies all the conditions in Definition 2.4.3.

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 2.4.7. *Suppose our nonstandard model is more saturated than the cardinality of X, then we can construct S_{ε}^{r} so that $X \subset S_{\varepsilon}^{r}$.*

Proof. Let $\mathscr{A} = \{U_i : i \leq K\}$ be the same object as in Theorem 2.4.6 and let $S_{\varepsilon}^{r} = \{s_i : i \leq K\}$ be a hyperfinite representation constructed from $\mathscr{A}$. 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

$$(S \in a) \wedge ((\forall s \in S)(\exists! U \in \mathscr{A})(s \in U)) \wedge (x \in S). \tag{2.4.4}$$

Consider the family $\mathscr{F} = \{\phi_x(S) | x \in X\}$, 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 $f(S_\varepsilon^r)$ 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 $\mathscr{F}$ simultaneously. This S_0 is the desired set. □

Let $(X, \mathscr{B}[X], P)$ be a Borel probability space satisfying the conditions of Theorem 2.4.6 and let S be an (ε, r)-hyperfinite representation of ${}^*X$. We now show that we can define an internal measure on $(S, \mathscr{I}(S))$ such that the resulting internal probability space is a good representation of $(X, \mathscr{B}[X], P)$. Similar theorems have been given assuming that X is merely Hausdorff [2]. Here we assume X is a σ-compact complete metric space satisfying Heine-Borel conditions and as a consequence we will obtain tighter control on the representation of $(X, \mathscr{B}[X], P)$.

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

Lemma 2.4.8 ([3, Thm 4.1])**.** *Let $(X, \mathscr{B}[X], \mu)$ be a σ-compact Borel probability space. Then* st *is measure preserving from $({}^*X, \overline{{}^*\mathscr{B}[X]}, \overline{{}^*\mu})$ to $(X, \mathscr{B}[X], \mu)$. That is, we have $\mu(E) = \overline{{}^*\mu}(\mathsf{st}^{-1}(E))$ for all $E \in \mathscr{B}[X]$.*

Proof. Let $E \in \mathscr{B}[X]$, $\varepsilon \in \mathbb{R}^+$ and choose K compact, U open with $K \subset E \subset U$ and $\mu(U) - \mu(K) < \varepsilon$. Note that ${}^*K \subset \mathsf{st}^{-1}(K) \subset \mathsf{st}^{-1}(E) \subset \mathsf{st}^{-1}(U) \subset {}^*U$, and ${}^*\overline{\mu}({}^*U) - {}^*\overline{\mu}({}^*K) < \varepsilon$. By Theorem 2.3.9, we have $\mathsf{st}^{-1}(E) \in \overline{{}^*\mathscr{B}[X]}$. 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 2.4.9. *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 $E \in \mathscr{B}[X]$, 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 $\mathscr{L}(\mathscr{A})$ denote the collection of universally Loeb measurable sets of the internal algebra $\mathscr{A}$.

Lemma 2.4.10. *Let X be a σ-compact metric space satisfying the Heine-Borel condition equipped with Borel σ-algebra $\mathscr{B}[X]$. Let S be a (ε,r)-hyperfinite representation of ${}^*X$ for some positive infinitesimal ε. Then for any $E \in \mathscr{B}[X]$ we have*

$$\mathsf{st}^{-1}(E) \in \mathscr{L}({}^*\mathscr{B}[X]) \text{ and } \mathsf{st}^{-1}(E) \cap S \in \mathscr{L}(\mathscr{I}(S)). \tag{2.4.5}$$

Proof. By Theorem 2.3.9, $\mathsf{st}^{-1}(E) \in \{A \cap \mathrm{NS}({}^*X) : A \in \mathscr{L}({}^*\mathscr{B}[X])\}$. As X is σ-compact, by Lemma 2.3.5, we have $\mathrm{NS}({}^*X) \in \mathscr{L}({}^*\mathscr{B}[X])$ hence $\mathsf{st}^{-1}(E) \in \mathscr{L}({}^*\mathscr{B}[X])$. Let P be any internal probability measure on $(S,\mathscr{I}(S))$ Let P' be an internal probability measure on $({}^*X, {}^*\mathscr{B}[X])$ with $P'(B) = P(B \cap S)$. As S is internal and $\mathsf{st}^{-1}(E)$ is universally Loeb measurable, we know that $\mathsf{st}^{-1}(E) \cap S \in \overline{{}^*\mathscr{B}[X]}^{P'}$ where $\overline{{}^*\mathscr{B}[X]}^{P'}$ denotes the Loeb σ-algebra of ${}^*\mathscr{B}[X]$ under P'. Fix any $\varepsilon > 0$. We can then find $A_i, A_o \in {}^*\mathscr{B}[X]$ such that $A_i \subset \mathsf{st}^{-1}(E) \cap S \subset A_o$ and $P'(A_o \setminus A_i) < \varepsilon$. We thus have

$$P'(A_o \setminus A_i) = P((A_o \setminus A_i) \cap S) = P((A_o \cap S) \setminus (A_i \cap S)) < \varepsilon. \tag{2.4.6}$$

As both $A_i, A_o \in {}^*\mathscr{B}[X]$, we know that $A_i \cap S, A_o \cap S \in \mathscr{I}(S)$. 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 $\mathsf{st}^{-1}(E) \cap S \in \mathscr{L}(\mathscr{I}(S))$. ☐

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

Theorem 2.4.11. *Let $(X, \mathscr{B}[X], P)$ be a Borel probability space where X is a σ-compact complete metric space satisfying the Heine-Borel condition, and let $({}^*X, {}^*\mathscr{B}[X], {}^*P)$ 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 $(S,\mathscr{I}(S))$*

1. *$P'(\{s\}) \approx {}^*P(B(s))$.*

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

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

Proof. Fix an infinitesimal $\varepsilon \in {}^*\mathbb{R}^+$ and an positive infinite number r. Let S be a (ε,r)-hyperfinite representation of ${}^*X$ and consider the hyperfinite measurable space $(S,\mathscr{I}(S))$. 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 $A \in \mathscr{I}(S)$, 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 $(S, \mathscr{I}(S))$.

As $\bigcup_{s\in S} B(s) \supset \mathrm{NS}({}^*X)$, by Lemma 2.4.8, 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 $E \in \mathscr{B}[X]$. As X is a σ-compact Borel probability space, by Lemma 2.4.8 and Lemma 2.4.10, we have $P(E) = \overline{{}^*P}(\mathsf{st}^{-1}(E))$. By Lemma 2.4.9, we then have

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \overline{{}^*P}(\bigcup\{B(s) : s \in \mathsf{st}^{-1}(E) \cap S\}). \tag{2.4.7}$$

Consider any set $A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathscr{I}(S)$, 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

$$\overline{{}^*P}(\bigcup\{B(s) : s \in \mathsf{st}^{-1}(E) \cap S\}) \leq \overline{{}^*P}(\bigcup_{s\in A_o} B(s)) = \mathsf{st}({}^*P(\bigcup_{s\in A_o} B(s))). \tag{2.4.8}$$

As $\bigcup_{s\in S} B(s) \supset \mathrm{NS}({}^*X)$, by Lemma 2.4.8, we have ${}^*P(\bigcup_{s\in S} B(s)) \approx 1$. Thus we have

$$\mathsf{st}({}^*P(\bigcup_{s\in A_o} B(s))) = \mathsf{st}(\frac{{}^*P(\bigcup_{s\in A_o} B(s))}{{}^*P(\bigcup_{s\in S} B(s))}) = \mathsf{st}(P'(A_o)) = \overline{P'}(A_o). \tag{2.4.9}$$

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

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \overline{{}^*P}(\bigcup\{B(s) : s \in \mathsf{st}^{-1}(E) \cap S\}) \leq \overline{P'}(A_o). \tag{2.4.10}$$

This means that

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) \leq \inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E) \cap S, A_o \in \mathscr{I}(S)\}. \tag{2.4.11}$$

By a similar argument, we have

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) \geq \sup\{\overline{P'}(A_i) : A_i \subset \mathsf{st}^{-1}(E) \cap S, A_i \in \mathscr{I}(S)\}. \tag{2.4.12}$$

By Lemma 2.4.10, we have $\mathsf{st}^{-1}(E)\cap S \in \mathscr{I}(S)_L$. Thus by the construction of Loeb measure, we have

$$\inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E)\cap S, A_o \in \mathscr{I}(S)\} \tag{2.4.13}$$

$$= \sup\{\overline{P'}(A_i) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\} \tag{2.4.14}$$

$$= \overline{P'}(\mathsf{st}^{-1}(E)\cap S). \tag{2.4.15}$$

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

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \inf\{\overline{P'}(A_o) : A_o \supset \mathsf{st}^{-1}(E)\cap S, A_o \in \mathscr{I}(S)\} \tag{2.4.16}$$

$$= \inf\{\mathsf{st}(\frac{{}^*P(\bigcup_{s\in A_o} B(s))}{{}^*P(\bigcup_{s\in S} B(s))}) : A_o \supset \mathsf{st}^{-1}(E)\cap S, A_o \in \mathscr{I}(S)\} \tag{2.4.17}$$

$$= \inf\{\overline{{}^*P}(\bigcup_{s\in A_o} B(s)) : A_o \supset \mathsf{st}^{-1}(E)\cap S, A_o \in \mathscr{I}(S)\}. \tag{2.4.18}$$

Similarly we have:

$$\overline{{}^*P}(\mathsf{st}^{-1}(E)) = \sup\{\overline{{}^*P}(\bigcup_{s\in A_i} B(s)) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\} \tag{2.4.19}$$

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 2.4.11. We no longer need to normalize the probability space when X is compact.

We conclude this section by giving an explicit application of Theorem 2.4.11 to Example 2.2.5.

Example 2.4.12. 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 $(\Omega, \mathscr{I}(\Omega))$ such that $\mu'(\{\omega\}) = \frac{1}{N}$ for every $\omega \in \Omega$.

Theorem 2.4.13. *For every Borel measurable set A, we have*

$$\mu(A) = \overline{\mu'}(\mathsf{st}^{-1}(A)\cap\Omega). \tag{2.4.20}$$

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 2.4.11, we have $\mu(A) = \overline{\mu'}(\mathsf{st}^{-1}(A)\cap\Omega)$, completing the proof. □

Chapter 3

Hyperfinite Representation of Standard Markov Processes

A Hyperfinite set is an infinite set with the same first-order logic properties as finite sets. A Hyperfinite stochastic process is a stochastic process with hyperfinite state space and time line. Thus, a hyperfinite stochastic process is a "continuous" stochastic process with the same first-order logic properties as discrete stochastic processes. There is a rich literature on studying hyperfinite stochastic process. In [1], Robert Anderson constructed a standard Brownian motion from a hyperfinite random walk, and the Itô stochastic integral from a hyperfinite sum. In [23], Jerome Keisler construct solutions of stochastic integral equations from solutions of hyperfinite difference equations. In this chapter, we extend Anderson's construction of the standard Brownian motion to all Markov processes satisfying certain regularity conditions. In particular, we will construct a hyperfinite Markov process for every standard Markov process such that their transition probability differ only by some infinitesimal. Unlike the construction in [23], our construction of hyperfinite Markov processes only depend on the transition probabilities of the original Markov processes.

In Section 3.1, we define hyperfinite Markov processes and investigate many of its properties. A hyperfinite Markov process is characterized by the following 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 $T=\{0,\delta t,\dots,K\}$ 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 very much like discrete-time Markov processes with finite state spaces. The Markov chain ergodic theorem for discrete-time Markov processes with finite state spaces can be proved using the "coupling" technique. Namely, for finite Markov processes, we can show that two i.i.d Markov processes 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 3.1.8. In Theorems 3.1.19 and 3.1.26, we establish the ergodic theorem for hyperfinite Markov processes.

In Section 3.2, 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 probabilities of $\{X'_t\}_{t\in T}$ deviate from the transition probabilities of $\{X_t\}_{t\geq 0}$ only by some infinitesimal. The hyperfinite Markov process $\{X'_t\}_{t\in T}$ is defined on some hyperfinite representation S of X. Note that the time line T of $\{X'_t\}_{t\in T}$ in this case is $\{1,2,\ldots,K\}$ for some infinite $K \in {}^*\mathbb{N}$. At each step, an infinitesimal difference between the internal transition probabilities of $\{X'_t\}_{t\in T}$ and the transition probabilities of $\{X_t\}$ is generated. As there are only countably many steps, the internal transition probabilities provide a reasonably well approximation for the transition probabilities of of $\{X_t\}_{t\in\mathbb{N}}$. We illustrate such result in Theorem 3.2.16.

In Section 3.3, we apply similar ideas developed in Section 3.2 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 3.2. When the time-line is continuous, the hyperfinite time-line T for the hyperfinite representation $\{X'_t\}_{t\in T}$ is $\{0,\delta t,2\delta t,\ldots,K\}$ where δt is some infinitesimal and K is some infinite number. As it takes hyperfinitely many infinitesimal steps to reach a non-infinitesimal time point, we need to make sure that the difference between the transition probabilities $\{X_t\}_{t\geq 0}$ and the internal transition probabilities $\{X'_t\}_{t\in T}$ generated in every step is so small such that the accumulated difference will remain infinitesimal. We establish this by using the internal induction principle (see Theorem 3.3.16). Unlike the construction of $\{X'_t\}_{t\in T}$ in Section 3.2, the construction of $\{X'_t\}_{t\in T}$ in Section 3.3 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 3.3.39.

3.1 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 3.1.1. 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 3.1.2. *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 3.1.1. Then there exists internal probability triple* $(\Omega, \mathscr{A}, P)$ *with an internal stochastic process* $\{X_t\}_{t \in T}$ *defined on* $(\Omega, \mathscr{A}, P)$ *such that*

$$P(X_0 = i_0, X_{\delta t} = i_{\delta t}, ... X_t = i_t) = v_{i_0} p_{i_0 i_{\delta t}} ... p_{i_{t-\delta t} i_t} \tag{3.1.1}$$

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 $\mathscr{A}$ be the set consisting of all internal subsets of Ω. We now define the internal measure P on $(\Omega, \mathscr{A})$. For every $\omega \in \Omega$, let

$$P(\omega) = v_{i_{\omega(0)}} p_{i_{\omega(0)} i_{\omega(1)}} \cdots p_{i_{\omega(K-\delta t)} i_{\omega(K)}}. \tag{3.1.2}$$

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

We use $(\Omega, \overline{\mathscr{A}}, \overline{P})$ to denote the Loeb extension of the internal probability triple $(\Omega, \mathscr{A}, P)$ in Theorem 3.1.2. 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 3.1.3. For any $i, j \in S$ and any $t \in T$, we define:

$$p_{ij}^{(t)} = \sum_{\omega \in \mathscr{M}} P(\{\omega\} | X_0 = i) \tag{3.1.3}$$

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

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 $A \in \mathscr{I}(S)$, 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 3.1.4. *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.

$$P(X_{k+\delta t} = j | X_k = i) = \frac{P(X_{k+\delta t} = j, X_k = i)}{P(X_k = i)} \tag{3.1.4}$$

$$= \frac{\sum_{i_0, i_{\delta t}, \ldots, i_{k-\delta t}} v_{i_0} p_{i_0 i_{\delta t}} \cdots p_{i_{k-\delta t} i} p_{ij}}{\sum_{i_0, i_{\delta t}, \ldots, i_{k-\delta t}} v_{i_0} p_{i_0 i_{\delta t}} \cdots p_{i_{k-\delta t} i}} \tag{3.1.5}$$

$$= p_{ij}. \tag{3.1.6}$$

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 $t \in T$, $p_i^{(t)}(.)$ is an internal probability measure on $(S, \mathscr{I}(S))$. We use $\overline{p}_i^{(t)}$ to denote the Loeb extension of this internal probability measure. For every $A \in \overline{\mathscr{I}(S)}$, 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 3.1.5. Let π be an internal probability measure on $(S, \mathscr{I}(S))$. 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 $A \in \mathscr{I}(S)$ 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 3.1.6. A hyperfinite Markov chain is said to be strong regular if for any $A \in \mathscr{I}(S)$, any $i, j \in \mathrm{NS}(S)$ and any non-infinitesimal $t \in T$ we have

$$(i \approx j) \implies (P_i(X_t \in A) \approx P_j(X_t \in A)). \tag{3.1.7}$$

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 3.1.7. 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 [1]). It can also be viewed as a hyperfinite Markov process. As the normal distributions with different means converge in total variation 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 3.1.8. *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 3.1.9. 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

$$\{q_{(i,j),(k,l)} = p_{ik}p_{jl}\}_{i,j,k,l\in S}. \tag{3.1.8}$$

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 3.1.10. *Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ and $\{Z_t\}_{t\in T}$ be the same as in Definition 3.1.9. Then for any $t \in T$,any $i,j \in S$ and any $A,B \in \mathscr{I}(S)$ 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 $A,B \in \mathscr{I}(S)$. We have

$$p_i^{(\delta t)}(A)p_j^{(\delta t)}(B) \tag{3.1.9}$$
$$= \sum_{(a,b)\in A\times B} p_i^{(\delta t)}(\{a\}) \times p_j^{(\delta t)}(\{b\}) \tag{3.1.10}$$
$$= \sum_{(a,b)\in A\times B} q_{(i,j)}^{(\delta t)}(\{(a,b)\}) \tag{3.1.11}$$
$$= q_{(i,j)}^{(\delta t)}(A\times B). \tag{3.1.12}$$

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 $A, B \in \mathscr{I}(S)$. We have

$$p_i^{(k+\delta t)}(A) \times p_j^{(k+\delta t)}(B) \tag{3.1.13}$$

$$= \sum_{s \in S} p_i^{(\delta t)}(\{s\}) p_s^{(k)}(A) \times \sum_{s' \in S} p_j^{(\delta t)}(\{s'\}) p_{s'}^{(k)}(B) \tag{3.1.14}$$

$$= \sum_{(s,s') \in S \times S} p_i^{(\delta t)}(\{s\}) p_j^{(\delta t)}(\{s'\}) p_s^{(k)}(A) p_{s'}^{(k)}(B) \tag{3.1.15}$$

By induction hypothesis, this equals to:

$$\sum_{(s,s') \in S \times S} q_{(i,j)}^{(\delta t)}(\{(s,s')\}) q_{(s,s')}^{(k)}(A \times B) = q_{(i,j)}^{(k+\delta t)}(A \times B). \tag{3.1.16}$$

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

Definition 3.1.11. 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 3.1.12. For any $A \in \mathscr{I}(S)$, 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 3.1.13. 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 3.1.9.

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 3.1.14. *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 3.1.11.*

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 standard result in [38].

Lemma 3.1.15 ([38, Lemma. 20])**.** *Consider a general hyperfinite Markov chain on a state space S, having a weakly stationary distribution* $\pi(.)$ *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 3.1.16. *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 $B_1\in\mathscr{F}$ 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 $\mathscr{F}$ denote the Loeb extension of the internal algebra $\mathscr{I}(S)$ 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

$$\overline{P}_x(\tau(A)=\infty,\eta<\infty) \tag{3.1.17}$$

$$=1-\overline{P}_x(\tau(A)=\infty,\eta=\infty)-\overline{P}_x(\tau(A)<\infty) \tag{3.1.18}$$

$$\geq 1-0-(1-\delta_1)=\delta_1. \tag{3.1.19}$$

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

$$\max\{t\in T:X_t\in B_2\}\geq\max\{t\in T:X_t\in B\} \tag{3.1.20}$$

establishing the claim. ☐

Claim 3.1.17. *Let B, l, δ be as in Claim 3.1.16. Let K' be the biggest hyperinteger such that $K'l \leq K$ where K is the last element in T. Let*

$$s = \max\{k \in {}^*\mathbb{N} : (1 \leq k \leq K') \wedge (X_{kl} \in B)\} \tag{3.1.21}$$

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

$$\sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) \gtrapprox \mathsf{st}(\pi(B)\delta). \tag{3.1.22}$$

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

$$\sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) = \sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y) P_y(s = 0, X_{(j-r)l} \notin A) \tag{3.1.23}$$

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

$$\sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y) P_y(s = 0, X_{(j-r)l} \notin A) \tag{3.1.24}$$

$$\geq \sum_{x \in S} \pi(\{x\}) \sum_{y \in B} P_x(X_{rl} = y)\delta \tag{3.1.25}$$

$$\approx \pi(B)\delta. \tag{3.1.26}$$

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 3.1.16, we have

$$\pi(A^c) \approx \sum_{x \in S} \pi(\{x\}) P_x(X_{jl} \in A^c) \tag{3.1.27}$$

$$= \sum_{x \in S} \pi(\{x\}) P_x(X_{jl} \notin A) \tag{3.1.28}$$

$$\geq \sum_{r=1}^{j} \sum_{x \in S} \pi(\{x\}) P_x(s = r, X_{jl} \notin A) \tag{3.1.29}$$

$$\geq \sum_{r=1}^{j} \mathsf{st}(\pi(B)\delta). \tag{3.1.30}$$

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 3.1.14.

proof of Theorem 3.1.14. 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 $U_{i_0}^{\frac{1}{n}} \times U_{i_0}^{\frac{1}{n}} \in \overline{\mathscr{I}(S)} \times \overline{\mathscr{I}(S)}$. By Definition 3.1.13, 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 3.1.15, 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 3.1.11, 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 3.1.14, 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 $A \in \mathscr{I}(S)$ and all $t \leq t_0$.

Lemma 3.1.18. *Let $\pi' = \pi \times \pi$. For any $A, B \in \mathscr{I}(S)$ and any $t \leq t_0$, we have $\pi'(A \times B) \approx \sum_{(i,j) \in S \times S} \pi'_{(i,j)} q_{(i,j)}^{(t)}(A \times B)$ where $q_{(i,j)}^{(t)}(A \times B)$ denotes the t-step transition probability from (i,j) to the set $A \times B$.*

Proof. Pick $A, B \in \mathscr{I}(S)$ and $t \leq t_0$. Then, by Definition 3.1.5 and Lemma 3.1.10, we have

$$\sum_{(i,j) \in S \times S} \pi'(\{(i,j)\}) q_{(i,j)}^{(t)}(A \times B) \tag{3.1.31}$$

$$= \sum_{(i,j) \in S \times S} \pi(\{i\}) \pi(\{j\}) p_i^{(t)}(A) p_j^{(t)}(B) \tag{3.1.32}$$

$$= (\sum_{i \in S} \pi(\{i\}) p_i^{(t)}(A))(\sum_{j \in S} \pi(\{j\}) p_j^{(t)}(B)) \tag{3.1.33}$$

$$\approx \pi(A)\pi(B) \tag{3.1.34}$$

$$= \pi(A \times B). \tag{3.1.35}$$

☐

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

Open Problem 2. *Does there exists a π' that fails to be a weakly stationary distribution of the product hyperfinite Markov process $\{Z_t\}_{t\in T}$?*

More generally, we ask the following question:

Open Problem 3. *Let P_1, P_2 be two internal probability measures on $(\Omega, \mathscr{A})$. Suppose $P_1(A) \approx P_2(A)$ for all $A \in \mathscr{A}$. Is it true that $(P_1 \times P_1)(B) \approx (P_2 \times P_2)(B)$ for all $B \in \mathscr{A} \times \mathscr{A}$?*

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

Theorem 3.1.19. *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 3.1.11, we know that for $\overline{\pi'}$ almost surely $(i, j) \in S \times S$ and any $n \in \mathbb{N}$ we have

$$\overline{P'}_{(i,j)}((\exists t \in \mathrm{NS}(T))(Z_t \in U_{i_0}^{\frac{1}{n}} \times U_{i_0}^{\frac{1}{n}})) = 1. \tag{3.1.36}$$

By Lemma 3.1.8, 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 $(i, j) \in S \times S$:

$$|\overline{P}_i(X_{t_1} \in A) - \overline{P}_j(X_{t_1} \in A)| \tag{3.1.37}$$
$$= |\overline{P'}_{(i,j)}(X_{t_1} \in A) - \overline{P'}_{(i,j)}(Y_{t_1} \in A)| \tag{3.1.38}$$
$$= |\overline{P'}_{(i,j)}((X_{t_1} \in A) \cap M^c) - \overline{P'}_{(i,j)}((Y_{t_1} \in A) \cap M^c)| \tag{3.1.39}$$
$$\leq \overline{P'}_{(i,j)}(M^c) = 0 \tag{3.1.40}$$

To see Eq. (3.1.39), 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 3.1.20. *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 $\mathscr{I}(S \times S)$. Fix some $n \in \mathbb{N}$. Let

$$F^n = \{(i, j) \in S \times S : |P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \leq \frac{1}{n}\}. \tag{3.1.41}$$

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 $\varepsilon \in \mathbb{R}$ such that $\mathsf{st}(\pi(E^n)) \leq 1 - \varepsilon$. As $F^n \subset E^n \times S$, we have

$$\pi'(F^n) = \pi(E^n) \times \pi(S) \leq 1 - \varepsilon \tag{3.1.42}$$

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 $\varepsilon_1, \varepsilon_2 \in \mathbb{R}$ and an internal set $D^n \subset E^n$ such that $\overline{\pi}(D^n) = \varepsilon_1$ and $\overline{\pi}(F_i^n) < 1 - \varepsilon_2$ 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

$$\overline{\pi'}(F^n) \leq \overline{\pi'}(F^n \cup A) = \overline{\pi}'(F^n \setminus A) + \overline{\pi'}(A). \tag{3.1.43}$$

Note that

$$\overline{\pi'}(F^n \setminus A) \leq \overline{\pi'}((E^n \setminus D^n) \times S) \leq \overline{\pi'}((S \setminus D^n) \times S) \leq 1 - \varepsilon_1 \tag{3.1.44}$$

$$\overline{\pi'}(A) = \mathsf{st}(\pi(A)) = \mathsf{st}(\sum_{i \in D^n} \pi(\{i\})\pi(F_i^n)) \leq \mathsf{st}(\sum_{i \in D^n} \pi(\{i\})(1 - \varepsilon_2)) = \varepsilon_1(1 - \varepsilon_2). \tag{3.1.45}$$

In conclusion, $\overline{\pi'}(F^n) = \overline{\pi'}(F^n \setminus A) + \overline{\pi'}(A) \leq (1 - \varepsilon_1) + \varepsilon_1(1 - \varepsilon_2) < 1$. 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

$$|P_i(X_{t_1} \in A) - \pi(A)| \approx |\sum_{j\in S} \pi(\{j\})(P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A))| \tag{3.1.46}$$

$$\leq \sum_{j\in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)|. \tag{3.1.47}$$

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 $\varepsilon \in \mathbb{R}^+$, we can find an internal $F_i' \subset F_i$ such that $\pi(F_i') > 1 - \varepsilon$. Now for $\overline{\pi}$ almost all i we have

$$\sum_{j\in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \tag{3.1.48}$$

$$= \sum_{j\in S\setminus F_i'} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| + \sum_{j\in F_i'} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \tag{3.1.49}$$

The first part of the last equation is less than ε and the second part is infinitesimal. Thus we have $\sum_{j\in S} \pi(\{j\})|P_i(X_{t_1} \in A) - P_j(X_{t_1} \in A)| \lessapprox \varepsilon$. 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 3.1.21. *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 3.1.14 and 3.1.19. □

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 3.1.22. *Let $\mathscr{L}(\mathscr{I}(S))$ denote the collection of all universally Loeb measurable sets (see Definition 2.3.8). Under the same assumptions of Theorem 3.1.19. For every $B \in \mathscr{L}(\mathscr{I}(S))$, 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 $\mathsf{st}^{-1}(E) \in \mathscr{L}(\mathscr{I}(S))$ for every $E \in \mathscr{B}[X]$.

We now show that we can actually obtain a stronger type of convergence than in Theorem 3.1.19 and Corollary 3.1.21.

Definition 3.1.23. Given two hyperfinite probability spaces $(S, \mathscr{I}(S), P_1)$ and $(S, \mathscr{I}(S), P_2)$, the total variation distance is defined to be

$$\| P_1(.) - P_2(.) \| = \sup_{A \in \mathscr{I}(S)} |P_1(A) - P_2(A)|. \tag{3.1.50}$$

Lemma 3.1.24. *We have*

$$\| P_1(.) - P_2(.) \| \geq \sup_{f: S \to {}^*[0,1]} |\sum_{i \in S} P_1(\{i\}) f(i) - \sum_{i \in S} P_2(\{i\}) f(i)|. \tag{3.1.51}$$

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 $\| p_x^{(t)}(.) - \pi(.) \|$. Just as standard Markov chains, we can show that the total variation distance is non-increasing.

Lemma 3.1.25. *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 $\| p_x^{(t_1)}(.) - \pi(.) \| \gtrapprox \| p_x^{(t_1+t_2)}(.) - \pi(.) \|$*

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

$$|p_x^{(t_1+t_2)}(A) - \pi(A)| \lessapprox \| p_x^{(t_1)}(.) - \pi(.) \| . \tag{3.1.52}$$

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 3.1.19 and Corollary 3.1.21.

Theorem 3.1.26. *Under the same hypotheses in Theorem 3.1.19. For $\overline{\pi}$ almost every $s \in \mathrm{NS}(S)$, the sequence $\{\sup_{B\in\mathscr{L}(\mathscr{I}(S))} |\overline{P}_s(X_t \in B) - \overline{\pi}(B)| : t \in \mathrm{NS}(T)\}$ converges to 0.*

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

$$\sup_{B\in\mathscr{L}(\mathscr{I}(S))} |\overline{P}_s(X_t \in B) - \overline{\pi}(B)| < \varepsilon. \tag{3.1.53}$$

Pick any real $\varepsilon > 0$, by Theorem 3.1.19, we know that for any infinite $t \leq t_0$ we have $\| p_s^{(t)}(.) - \pi(.) \| < \frac{\varepsilon}{2}$. By underspill, there exist a $t_1 \in \mathrm{NS}(T)$ such that $\| p_s^{(t_1)}(.) - \pi(.) \| < \frac{\varepsilon}{2}$. Fix any $t_2 \geq t_1$. Then by Lemma 3.1.25 we have $\| p_s^{(t_2)}(.) - \pi(.) \| < \varepsilon$. Now fix any internal set $A \subset S$. By the definition of total variation distance, we have $|P_s(X_{t_2} \in A) - \pi(A)| < \varepsilon$. This implies that $|\overline{P}_s(X_{t_2} \in A) - \overline{\pi}(A)| \leq \varepsilon$ for all $A \in \mathscr{I}(S)$. For external $B \in \mathscr{L}(\mathscr{I}(S))$, we have

$$\overline{P}_s(X_{t_2} \in B) = \sup\{\overline{P}_s(X_{t_2} \in A_i) : A_i \subset B, A_i \in \mathscr{I}(S)\} \tag{3.1.54}$$

$$\overline{\pi}(B) = \sup\{\overline{\pi}(A_i) : A_i \subset B, A_i \in \mathscr{I}(S)\} \tag{3.1.55}$$

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

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

$$\lim_{t\to\infty}\{\sup_{E\in\mathscr{B}[X]} |\overline{P}_x^{(t)}(\mathsf{st}^{-1}(E)) - \overline{\pi}(\mathsf{st}^{-1}(E))| : t \in \mathrm{NS}(T)\} = 0. \tag{3.1.56}$$

Note that the statement of Theorem 3.1.26 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.

3.2 Hyperfinite Representation for Discrete-time Markov Processes

As one can see from Section 3.1, 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.

3.2.1 General properties of the transition probability

Consider a Markov chain $\{X_t\}_{t\geq 0}$ on $(X,\mathscr{B}[X])$ 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

$$\{P_x^{(t)}(A) : x \in X, t \in \mathbb{R}^+, A \in \mathscr{B}[X]\}. \tag{3.2.1}$$

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 $P_x^{(t)}(.)$ is a probability measure on $(X,\mathscr{B}[X])$. It is sometimes desirable to treat the transition probability as a function of three variables. Namely, we define a function $g : X \times \mathbb{R}^+ \times \mathscr{B}[X] \mapsto [0,1]$ 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 ${}^*X \times {}^*\mathbb{R}^+ \times {}^*\mathscr{B}[X]$ to ${}^*[0,1]$.

Lemma 3.2.1. *For any given $x \in {}^*X$, any $t \in {}^*\mathbb{R}^+$, ${}^*g(x,t,.)$ is an internal finitely-additive probability measure on $({}^*X, {}^*\mathscr{B}[X])$.*

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

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

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

□

Recall that for every fixed $A \in \mathscr{B}[X]$ 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 3.2.2. *For each fixed $A \in {}^*\mathscr{B}[X]$ 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 $\forall A \in \mathscr{B}[X]\ \forall t \in \mathbb{R}^+\ \forall B \in \mathscr{B}[[0,1]]\ \{x : g(x,t,A) \in B\} \in \mathscr{B}[X]$. By the transfer principle, we get the desired result. □

For every $x \in {}^*X$ and $t \in {}^*\mathbb{R}^+$, we use $\overline{{}^*P_x^{(t)}}(.)$ or $\overline{{}^*g}(x,t,.)$ 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 3.2.3. For any $A,B \in \mathscr{B}[X]$, 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 $A \in \mathscr{B}[X]$ it is easy to see that $f_x^{(k_1,k_2)}(A,.)$ is a probability measure on $(X,\mathscr{B}[X])$ 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 $X \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathscr{B}[X] \times \mathscr{B}[X]$ to $[0,1]$. By the transfer principle, we know that ${}^*f$ is an internal function from ${}^*X \times {}^*\mathbb{R}^+ \times {}^*\mathbb{R}^+ \times {}^*\mathscr{B}[X] \times {}^*\mathscr{B}[X]$ to ${}^*[0,1]$. Moreover, ${}^*f_x^{(k_1,k_2)}(A,.)$ is an internal probability measure on $({}^*X, {}^*\mathscr{B}[X])$ provided that ${}^*g(x,k_1,A) > 0$.

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

Lemma 3.2.4. *Consider any $k_1,k_2 \in \mathbb{R}^+$, any $x \in X$ and any two sets $A,B \in \mathscr{B}[X]$ such that $g(x,k_1,A) > 0$. If there exists an $\varepsilon > 0$ such that for any two points $x_1,x_2 \in A$ we have $|g(x_1,k_2,B) - g(x_2,k_2,B)| \leq \varepsilon$, then for any point $y \in A$ we have $|g(y,k_2,B) - f_x^{(k_1,k_2)}(A,B)| \leq \varepsilon$.*

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

$$f_x^{(k_1,k_2)}(A,B) = \frac{P_x(X_{k_1+k_2} \in B, X_{k_1} \in A)}{P_x(X_{k_1} \in A)} = \frac{\int_A g(s,k_2,B)g(x,k_1,\mathrm{d}s)}{g(x,k_1,A)}. \tag{3.2.2}$$

For any $y \in A$, we have

$$|g(y,k_2,B) - f_x^{(k_1,k_2)}(A,B)| = \frac{\int_A |g(y,k_2,B) - g(s,k_2,B)| g(x,k_1.ds)}{g(x,k_1,A)}. \tag{3.2.3}$$

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

$$\frac{\int_A |g(y,k_2,B) - g(s,k_2,B)| g(x,k_1.ds)}{g(x,k_1,A)} \leq \frac{\varepsilon \cdot g(x,k_1,A)}{g(x,k_1,A)} = \varepsilon. \tag{3.2.4}$$

□

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 3.2.4, we obtain the following lemma

Lemma 3.2.5. *Consider any $k_1,k_2 \in {}^*\mathbb{R}^+$, any $x \in {}^*X$ and any two internal sets $A,B \in {}^*\mathscr{B}[X]$ such that $g(x,k_1,A) > 0$. If there exists a positive $\varepsilon \in {}^*\mathbb{R}$ such that for any two points $x_1,x_2 \in A$ we have $|{}^*g(x_1,k_2,B) - {}^*g(x_2,k_2,B)| \leq \varepsilon$, then for any point $y \in A$ we have $|{}^*g(y,k_2,B) - {}^*f_x^{(k_1,k_2)}(A,B)| \leq \varepsilon$.*

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 3.2.4 and 3.2.5 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 $f_x^{(k_1,k_2)}(A,B)$.

3.2.2 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 $g : X \times \mathbb{N} \times \mathscr{B}[X] \mapsto [0,1]$ in a same way as in last section. The nonstandard extension ${}^*g$ is an internal function from ${}^*X \times {}^*\mathbb{N} \times {}^*\mathscr{B}[X]$ 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 $\varepsilon > 0$ there exists $\delta > 0$ such that

$$(\forall x_1 \in X)(|x_1 - x| < \delta \implies ((\forall A \in \mathscr{B}[X])|P_{x_1}(A) - P_x(A)| < \varepsilon)). \tag{3.2.5}$$

We quote the following lemma regarding total variation distance. This lemma is the "standard counterpart" of Lemma 3.1.24.

Lemma 3.2.6 ([38])**.** *Let ν_1, ν_2 be two different probability measures on some space $(X, \mathscr{F})$ 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 3.2.6 is the following result which can be viewed as a discrete-time counterpart of Lemma 3.1.25.

Lemma 3.2.7. *Consider the discrete-time Markov process $\{X_t\}_{t\in\mathbb{N}}$ with state space X. For every $\varepsilon > 0$, every $x_1, x_2 \in X$ and every positive $k \in \mathbb{N}$ we have*

$$((\forall A \in \mathscr{B}[X])(|P_{x_1}^{(k)}(A) - P_{x_2}^{(k)}(A)| \leq \varepsilon)) \implies ((\forall A \in \mathscr{B}[X])(|P_{x_1}^{(k+1)}(A) - P_{x_2}^{(k+1)}(A)| \leq \varepsilon)). \tag{3.2.6}$$

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

$$\sup_{A\in\mathscr{B}[X]} \{|P_{x_1}^{(k+1)}(A) - P_{x_2}^{(k+1)}(A)|\} \tag{3.2.7}$$

$$= \sup_{A\in\mathscr{B}[X]} \{|\int_{y\in X} P_y(A) P_{x_1}^{(k)}(dy) - \int_{y\in X} P_y(A) P_{x_2}^{(k)}(dy)|\} \tag{3.2.8}$$

$$\leq \parallel P_{x_1}^{(k)}(.) - P_{x_2}^{(k)}(.) \parallel \leq \varepsilon. \tag{3.2.9}$$

Thus we have proved the result. □

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

Lemma 3.2.8. *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 $A \in {}^*\mathscr{B}[X]$ 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

$$(\forall x \in {}^*X)(|x - x_0| < \delta \implies ((\forall A \in {}^*\mathscr{B}[X])|{}^*g(x, 1, A) - {}^*g(x_0, 1, A)| < \varepsilon)) \tag{3.2.10}$$

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 $A \in {}^*\mathscr{B}[X]$.

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

$$(\forall A \in {}^*\mathscr{B}[X])(|{}^*g(x_1,1,A) - {}^*g(x_2,1,A)| < \varepsilon). \tag{3.2.11}$$

By the transfer of Lemma 3.2.7, we know that for every $k \in \mathbb{N}$ we have

$$(\forall A \in {}^*\mathscr{B}[X])(|{}^*g(x_1,k,A) - {}^*g(x_2,k,A)| < \varepsilon). \tag{3.2.12}$$

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 $B(s) \in {}^*\mathscr{B}[X]$ 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 $({}^*X, {}^*\mathscr{B}[X])$. 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 3.2.9. For $i \in \{0,1\}$, let $g'(x,i,A) : \bigcup_{s\in S} B(s) \times {}^*\mathscr{B}[X] \to {}^*[0,1]$ be given by:

$$g'(x,i,A) = {}^*g(x,i,A \cap \bigcup_{s\in S} B(s)) + \delta_x(A)\,{}^*g(x,i,{}^*X \setminus \bigcup_{s\in S} B(s)). \tag{3.2.13}$$

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 $A \in {}^*\mathscr{B}[X]$, 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 3.2.10. *Let $\mathscr{B}[\bigcup_{s\in S} B(s)] = \{A \cap \bigcup_{s\in S} B(s) : A \in {}^*\mathscr{B}[X]\}$. Then for any $x \in \bigcup_{s\in S} B(s)$, the triple $(\bigcup_{s\in S} B(s), \mathscr{B}[\bigcup_{s\in S} B(s)], g'(x,1,.))$ 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 $(\bigcup_{s\in S} B(s), \mathscr{B}[\bigcup_{s\in S} B(s)])$.

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 $A,B \in \mathscr{B}[\bigcup_{s\in S} B(s)]$, we have:

$$g'(x,1,A\cup B) \tag{3.2.14}$$

$$= {}^*g(x,1,A\cup B)+\delta_x(A\cup B)^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s)) \tag{3.2.15}$$

$$= {}^*g(x,1,A)+\delta_x(A)^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s))+{}^*g(x,1,B)+\delta_x(B)^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s)) \tag{3.2.16}$$

$$= g'(x,1,A)+g'(x,1,B). \tag{3.2.17}$$

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 3.2.11. *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 3.2.8, we know that ${}^*g(x,t,A)\approx {}^*g(x_0,t,A)$ for every $A\in {}^*\mathscr{B}[X]$. 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 $(S,\mathscr{I}(S))$ 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 $(S,\mathscr{I}(S))$ with "one-step" transition probability $\{G_{ij}\}_{i,j\in S}$. We first verify that $G_i(.)$ is an internal probability measure on $(S,\mathscr{I}(S))$ for every $i\in S$.

Lemma 3.2.12. *For every $i\in S$, $G_i(.)$ and $G^{(0)}_i(.)$ are internal probability measure on $(S,\mathscr{I}(S))$.*

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

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

$$G_i(\emptyset) = g'(i,1,\emptyset) = 0 \tag{3.2.18}$$

$$G_i(S) = g'(i,1,\bigcup_{s\in S} B(s)) = {}^*g(i,1,\bigcup_{s\in S} B(s)) + \delta_i(\bigcup_{s\in S} B(s))^*g(i,1,{}^*X\setminus\bigcup_{s\in S} B(s)) = 1. \tag{3.2.19}$$

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 $G_i(A\cup B) = \sum_{j\in A\cup B} G_{ij} = G_i(A)+G_i(B)$. □

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

Lemma 3.2.13. *For any $i\in S$ and any $t\in\mathbb{N}$, $G_i^{(t)}(.)$ is an internal probability measure on $(S,\mathscr{I}(S))$.*

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

For t equals to 0 or 1, we already have the results by Lemma 3.2.12.

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 $A\in\mathscr{I}(S)$ 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 $A,B\in\mathscr{I}(S)$. We have:

$$G_i^{(t_0+1)}(A\cup B) = \sum_{j\in S} G_{ij}(G_j^{(t_0)}(A)+G_j^{(t_0)}(B)) = G_j^{(t_0+1)}(A)+G_j^{(t_0+1)}(B). \tag{3.2.20}$$

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

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

Theorem 3.2.14. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ satisfies (DSF). Then for any $n\in\mathbb{N}$, any $x\in \mathrm{NS}(S)$ and any $A\in {}^*\mathscr{B}[X]$, ${}^*g(x,n,\bigcup_{s\in A\cap S} B(s))\approx G_x^{(n)}(A\cap S)$.*

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 $A \in {}^*\mathscr{B}[X]$. We have

$$
\begin{aligned}
& G_x(A\cap S) && (3.2.21)\\
&= g'(x,1,\bigcup_{s\in A\cap S} B(s)) && (3.2.22)\\
&= {}^*g(x,1,\bigcup_{s\in A\cap S} B(s)) + \delta_x(\bigcup_{s\in A\cap S} B(s)){}^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s)) && (3.2.23)\\
&\approx {}^*g(x,1,\bigcup_{s\in A\cap S} B(s)) && (3.2.24)
\end{aligned}
$$

where the last $\approx$ follows from Lemma 3.2.11.

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

$$
\begin{aligned}
& {}^*g(x,k+1,\bigcup_{s'\in A\cap S} B(s')) && (3.2.25)\\
&= (\sum_{s\in S} {}^*g(x,1,B(s)){}^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S} B(s')) && (3.2.26)\\
&+ {}^*g(x,1,{}^*X\setminus \bigcup_{s\in S} B(s)){}^*f_x^{(k)}({}^*X\setminus \bigcup_{s\in S} B(s),\bigcup_{s'\in A\cap S} B(s')) && (3.2.27)\\
&\approx \sum_{s\in S} {}^*g(x,1,B(s)){}^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S} B(s')). && (3.2.28)
\end{aligned}
$$

where the last $\approx$ follows from Lemma 3.2.11.

By Lemmas 3.2.5 and 3.2.8, 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

$$
\sum_{s\in S} {}^*g(x,1,B(s)){}^*f_x^{(k)}(B(s),\bigcup_{s'\in A\cap S} B(s')) \approx \sum_{s\in S} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')). \qquad (3.2.29)
$$

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 $\varepsilon \in \mathbb{R}$. By Lemma 3.2.11, we can pick an internal set $M \subset \mathrm{NS}(S)$ such that ${}^*g(x,1,\bigcup_{s\in M} B(s)) > 1-\varepsilon$. We then have

$$
\begin{aligned}
& \sum_{s\in S} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) && (3.2.30)\\
&= \sum_{s\in M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) + \sum_{s\in S\setminus M} {}^*g(x,1,B(s)){}^*g(s,k,\bigcup_{s'\in A\cap S} B(s')). && (3.2.31)
\end{aligned}
$$

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 2.1.20 we have

$$\sum_{s\in M} {}^*g(x,1,B(s))^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \approx \sum_{s\in M} {}^*g(x,1,B(s))G_s^{(k)}(A\cap S). \tag{3.2.32}$$

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

$$|\sum_{s\in M} {}^*g(x,1,B(s))G_s^{(k)}(A\cap S) - \sum_{s\in M} g'(x,1,B(s))G_s^{(k)}(A\cap S)| \tag{3.2.33}$$

$$= |({}^*g(x,1,B(s_0)) - g'(x,1,B(s_0)))G_{s_0}^{(k)}(A\cap S)| \tag{3.2.34}$$

$$= |{}^*g(x,1,{}^*X\setminus\bigcup_{s\in S} B(s))G_{s_0}^{(k)}(A\cap S)| \approx 0 \tag{3.2.35}$$

where the last $\approx$ follows from Lemma 3.2.11. Thus, by Eq. (3.2.32), we have

$$\sum_{s\in M} {}^*g(x,1,B(s))^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) \tag{3.2.36}$$

$$\approx \sum_{s\in M} g'(x,1,B(s))G_s^{(k)}(A\cap S) \tag{3.2.37}$$

$$= \sum_{s\in M} G_x(\{s\})G_s^{(k)}(A\cap S). \tag{3.2.38}$$

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

$$\sum_{s\in S\setminus M} {}^*g(x,1,B(s))^*g(s,k,\bigcup_{s'\in A\cap S} B(s')) < \varepsilon. \tag{3.2.39}$$

On the other hand, we have

$$\sum_{s\in S\setminus M} G_x(\{s\})G_s^{(k)}(A\cap S) \tag{3.2.40}$$

$$= \sum_{s\in S\setminus M} g'(x,1,B(s))G_s^{(k)}(A\cap S) \tag{3.2.41}$$

$$\leq \sum_{s\in S\setminus M} g'(x,1,B(s)) \tag{3.2.42}$$

$$\leq {}^*g(x,1,\bigcup_{s\in S\setminus M} B(s)) + {}^*g(x,1,{}^*X\setminus\bigcup_{s\in S} B(s)) \tag{3.2.43}$$

$$\approx {}^*g(x,1,\bigcup_{s\in S\setminus M} B(s)) < \varepsilon \tag{3.2.44}$$

where the second last $\approx$ follows from Lemma 3.2.11.

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

$$|{}^*g(x,k+1,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(k+1)}(A\cap S)| \lessapprox \varepsilon \tag{3.2.45}$$

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. □

The following lemma is a slight generalization of [3, Thm 4.1].

Lemma 3.2.15. *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 $\varepsilon\in\mathbb{R}$, as $g(x_0,n,.)$ is a Radon measure, we can find K compact, U open with $K\subset E\subset U$ such that $g(x_0,n,U)-g(x_0,n,K)<\frac{\varepsilon}{2}$. By the transfer principle, we know that ${}^*g(x_0,n,{}^*U)-{}^*g(x_0,n,{}^*K)<\varepsilon/2$. 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 ${}^*g(x,n,{}^*U)-{}^*g(x,n,{}^*K)<\varepsilon$. 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 $|{}^*g(x,n,{}^*E)-\overline{{}^*g}(x,n,\mathsf{st}^{-1}(E))|<\varepsilon$. 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 3.2.16. *Suppose* $\{X_t\}_{t\in\mathbb{N}}$ *satisfies (DSF). Then for any* $s \in \mathrm{NS}(S)$*, any* $n \in \mathbb{N}$ *and any* $E \in \mathscr{B}[X]$, $P^{(n)}_{\mathsf{st}(s)}(E) = \overline{G}^{(n)}_s(\mathsf{st}^{-1}(E) \cap S)$.

Proof. Fix any $s \in \mathrm{NS}(S)$, any $n \in \mathbb{N}$ and any Borel set E. By Lemma 3.2.15, 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. (2.4.19), we have

$$\overline{{}^*g}(s,n,st^{-1}(E)) = \sup\{\overline{{}^*g}(s,n,\bigcup_{s\in A_i} B(s)) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\}. \tag{3.2.46}$$

By Theorem 3.2.14, we have $\overline{{}^*g}(s,n,\bigcup_{s\in A_i} B(s)) = \overline{G}^{(n)}_s(A_i)$. Thus we have

$$\overline{{}^*g}(s,n,\mathsf{st}^{-1}(E)) = \sup\{\overline{G}^{(n)}_s(A_i) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\} = \overline{G}^{(n)}_s(\mathsf{st}^{-1}(E)\cap S). \tag{3.2.47}$$

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.

3.3 Hyperfinite Representation for Continuous-time Markov Processes

In Section 3.2.2, 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 3.2 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

$$\{P^{(t)}_x(A) : x \in X, t \in \mathbb{R}^+, A \in \mathscr{B}[X]\}. \tag{3.3.1}$$

When we view the transition probability as a function of three variables, we again use $g(x,t,A)$ to denote the transition probability $P^{(t)}_x(A)$. We have already established some general properties regarding the transition probability $g(x,t,A)$ in Section 3.2.1. We recall some important definitions are results here.

Definition 3.3.1. For any $A, B \in \mathscr{B}[X]$, any $k_1, k_2 \in \mathbb{R}^+$ and any $x \in X$, let $f^{(k_1,k_2)}_x(A,B)$ be $P_x(X_{k_1+k_2} \in B | X_{k_1} \in A)$ when $P^{(k_1)}_x(A) > 0$ and let $f^{(k_1,k_2)}_x(A,B) = 1$ 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 $A \in \mathscr{B}[X]$, we have

$$g(x, k_1 + k_2, A) = \sum_{n \in \mathbb{N}} g(x, k_1, A_n) f_x^{(k_1, k_2)}(A_n, A). \tag{3.3.2}$$

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 3.2.1, we are interested in the relation between the nonstandard extensions of g and f. Recall Lemma 3.2.5 from Section 3.2.1.

Lemma 3.3.2. *Consider any $k_1, k_2 \in {}^*\mathbb{R}^+$, any $x \in {}^*X$ and any two sets $A, B \in {}^*\mathscr{B}[X]$ such that $g(x, k_1, A) > 0$. If there exists a positive $\varepsilon \in {}^*\mathbb{R}$ such that for any two points $x_1, x_2 \in A$ we have $|{}^*g(x_1, k_2, B) - {}^*g(x_2, k_2, B)| \leq \varepsilon$, then for any point $y \in A$ we have $|{}^*g(y, k_2, B) - {}^*f_x^{(k_1, k_2)}(A, B)| \leq \varepsilon$.*

Let the hyperfinite time line $T = \{\delta t, \ldots, K\}$ as in Section 3.1. 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 3.2.2, 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.

3.3.1 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 $\varepsilon_0 \frac{t}{\delta t} \approx 0$ 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 start by making the following assumption:

Condition VD. The Markov chain $\{X_t\}_{t \geq 0}$ is said to *vanish in distance* if for all $t \geq 0$ and all compact $K \subset X$ we have:

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

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

An alternative but stronger assumption is

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

$$(\forall \varepsilon > 0)(\exists r > 0)(\forall x \in X)(\forall A \in \mathscr{B}[X])(d(x,A) > r \implies g(x,t,A) < \varepsilon). \tag{3.3.3}$$

It is easy to see that (SVD) implies (VD).

Example 3.3.3. The Ornstein-Uhlenbeck is a continuous time stochastic process $\{X_t\}_{t\geq 0}$ satisfies the stochastic differential equation:

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t. \tag{3.3.4}$$

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 (VD) but not (SVD).

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

$$\{x \in {}^*X : {}^*d(x,x_0) \leq r\} \tag{3.3.5}$$

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

Theorem 3.3.4. *Suppose (VD) holds. For every positive $\varepsilon \in {}^*\mathbb{R}$, there exists an open ball $U(a,r)$ centered at some standard point a with radius r such that:*

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

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

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

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

$$(\forall x \in {}^*K_n)(\forall A \in {}^*\mathscr{B}[X])({}^*d(x,A) > r \implies {}^*g(x,\delta t,A) < \varepsilon). \tag{3.3.6}$$

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

$$(\forall x \in \bigcup_{n\in\mathbb{N}} {}^*K_n)(\forall A \in {}^*\mathscr{B}[X])({}^*d(x,A) > r_{\delta t} \implies {}^*g(x,\delta t,A) < \varepsilon). \tag{3.3.7}$$

Claim 3.3.5. *For every $n \in \mathbb{N}$, the formula $\phi_n(r)$*

$$(\forall x \in {}^*X)({}^*d(x,{}^*K_n) > r \implies ((\forall t \in T)({}^*g(x,t,{}^*K_n) < \varepsilon))). \tag{3.3.8}$$

is satisfiable.

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

$$(\forall x \in {}^*X)({}^*d(x,{}^*K_n) > r \implies {}^*g(x,t,{}^*K_n) < \varepsilon). \tag{3.3.9}$$

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

$$h(t) = \min\{r \in {}^*\mathbb{R}^+ : (\forall x \in {}^*X)({}^*d(x,{}^*K_n) > r \implies {}^*g(x,t,{}^*K_n) < \varepsilon)\} \tag{3.3.10}$$

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

$$(\forall x \in {}^*X)(\forall n \in \mathbb{N})({}^*d(x,{}^*K_n) > r' \implies ((\forall t \in T)({}^*g(x,t,{}^*K_n) < \varepsilon))). \tag{3.3.11}$$

Consider any near-standard internal set A.

Claim 3.3.6. *There exists $n \in \mathbb{N}$ such that $A \subset {}^*K_n$.*

Proof. Suppose not. Then $\mathscr{M}_n = \{a \in A : a \notin {}^*K_n\}$ 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 $\bigcap_{n\in\mathbb{N}} \mathscr{M}_n \neq \emptyset$. Hence there exists $a \in A$ such that $a \notin \bigcup_{n\in\mathbb{N}} {}^*K_n$. By Theorem 2.1.28, 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 $A \in {}^*\mathscr{B}[X]$ we have

$$((\forall n \in \mathbb{N})({}^*d(x, {}^*K_n) > r')) \implies ((\forall t \in T)({}^*g(x,t,A) < \varepsilon)). \tag{3.3.12}$$

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. (3.3.7), we have

$$(\forall x \in \mathrm{NS}({}^*X))({}^*g(x, \delta t, {}^*X \setminus \overline{U}(a,r)) < \varepsilon). \tag{3.3.13}$$

Fix any $y \in {}^*X \setminus \overline{U}(a,r)$ and any near-standard $A \in {}^*\mathscr{B}[X]$. By the choice of r, we know that ${}^*d(y, {}^*K_n) > r'$ for all $n \in \mathbb{N}$. Thus, by Eq. (3.3.12) we have ${}^*g(y,t,A) < \varepsilon$ 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 3.3.4. We fix a_0 and r_1 for the remainder of this section.

Lemma 3.3.7. *Suppose (VD) holds. There exists a positive infinite $r_0 > 2r_1$ such that*

$$(\forall y \in \overline{U}(a_0, 2r_1))({}^*g(y, \delta t, {}^*X \setminus \overline{U}(a_0, r_0)) < \varepsilon_0). \tag{3.3.14}$$

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 (VD). Note that we can always pick r_0 to be bigger than $2r_1$. □

We will see how do we use Lemma 3.3.7 in Theorem 3.3.16. We now fix r_0 for the remainder of this section. An immediate consequence of Theorem 3.3.4 and Lemma 3.3.7 is:

Lemma 3.3.8. *Suppose (VD) holds. For any $x \in X$, any $t \in T$, any near-standard internal set $A \subset {}^*X$ we have ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, 2r_0), A) < 2\varepsilon_0$.*

Proof. Fix a $x \in {}^*X$, a near-standard internal set A and some $t \in T$. By Theorem 3.3.4, we know that $(\forall y \in {}^*X \setminus \overline{U}(a_0, 2r_0))({}^*g(y,t,A) < \varepsilon_0)$. This means that for any $y_1, y_2 \in {}^*X \setminus \overline{U}(a_0, 2r_0)$ we have

$|^*g(y_1,t,A) - {}^*g(y_2,t,A)| < \varepsilon_0$. By Lemma 3.2.5, we know that for any $y \in {}^*X \setminus \overline{U}(a_0, 2r_0)$ we have

$$|^*g(y,t,A) - {}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0,2r_0),A)| < \varepsilon_0 \tag{3.3.15}$$

which then implies that ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0,2r_0),A) < 2\varepsilon_0$. □

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 assume the following condition on the Markov chain $\{X_t\}_{t\geq 0}$. We shall use this condition to control the diameter of each $B(s)$ for $s \in S$.

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 $\varepsilon > 0$ there exists $\delta > 0$ such that

$$(\forall x \in X)(\exists \delta > 0)((\forall y \in X)(|y - x| < \delta \implies (\forall A \in \mathscr{B}[X])|P_y^{(t)}(A) - P_x^{(t)}(A)| < \varepsilon)). \tag{3.3.16}$$

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 $\varepsilon \in {}^*\mathbb{R}^+$ and every $x \in {}^*X$ there exists $\delta \in {}^*\mathbb{R}^+$ such that:

$$((\forall y \in {}^*X)(|y - x| < \delta \implies (\forall A \in {}^*\mathscr{B}[X])|^*g(y,t,A) - {}^*g(x,t,A)| < \varepsilon)). \tag{3.3.17}$$

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 3.1.25. The proof is identical to Lemma 3.2.7 hence omitted.

Lemma 3.3.9. *Consider a standard Markov process with transition probability measure $P_x^{(t)}(.)$, then for every $\varepsilon \in \mathbb{R}^+$, every $x_1, x_2 \in X$, every $t_1, t_2 \in \mathbb{R}^+$ and every $A \in \mathscr{B}[X]$ we have*

$$(|P_{x_1}^{(t_1)}(A) - P_{x_2}^{(t_1)}(A)| \leq \varepsilon \implies |P_{x_1}^{(t_1+t_2)}(A) - P_{x_2}^{(t_1+t_2)}(A)| \leq \varepsilon). \tag{3.3.18}$$

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

$$((|^*P_{x_1}^{(t_1)}(A) - {}^*P_{x_2}^{(t_1)}(A)| \leq \varepsilon) \implies (|^*P_{x_1}^{(t_1+t_2)}(A) - {}^*P_{x_2}^{(t_1+t_2)}(A)| \leq \varepsilon)). \tag{3.3.19}$$

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 3.3.10. *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 $|{}^*g(x_1,t,A) - {}^*g(x_2,t,A)| < \varepsilon_0$ for all $A \in {}^*\mathscr{B}[X]$ 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:

$$(\forall y \in {}^*X)(|y - x| < \delta_x \implies (\forall A \in {}^*\mathscr{B}[X])|{}^*g(x, \delta t, A) - {}^*g(y, \delta t, A)| < \frac{\varepsilon_0}{2}). \tag{3.3.20}$$

The internal collection $\mathscr{L} = \{U(x, \frac{\delta_x}{2}) : x \in \overline{U}(a_0, 2r_0)\}$ 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 $\mathscr{L}$ that covers $\overline{U}(a_0, 2r_0)$. Denote this hyperfinite subcover by $\mathscr{F} = \{B(x_i, \frac{\delta_{x_i}}{2}) : i \leq N\}$ 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 $|x - y| < \delta_0$. 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 $(\forall A \in {}^*\mathscr{B}[X])(|{}^*g(x, \delta t, A) - {}^*g(y, \delta t, A)| < \varepsilon_0)$. By Eq. (3.3.19), we know that $(\forall A \in {}^*\mathscr{B}[X])(\forall t \in T|{}^*g(x,t,A) - {}^*g(y,t,A)| < \varepsilon_0)$, 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 3.3.11. *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 $A \in {}^*\mathscr{B}[X]$ and any $t \in T^+$ we have $|{}^*g(x_1,t,A) - {}^*g(x_2,t,A)| < \varepsilon_0$*

An immediate consequence of the above lemma is:

Lemma 3.3.12. *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 $A \in {}^*\mathscr{B}[X]$ and any $t \in T^+$ we have $|{}^*g(y,t,A) - {}^*f_x^{(t)}(B(s), A)| < \varepsilon_0$.*

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 3.2.4 to Theorem 3.3.11. □

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 $\{X_t\}_{t\geq 0}$.

3.3.2 Construction of Hyperfinite Markov Processes

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 3.2.9.

Definition 3.3.13. Let $g'(x, \delta t, A) : \bigcup_{s\in S} B(s) \times {}^*\mathscr{B}[X] \to {}^*[0,1]$ be given by:

$$g'(x, \delta t, A) = {}^*g(x, \delta t, A \cap \bigcup_{s\in S} B(s)) + \delta_x(A){}^*g(x, \delta t, {}^*X \setminus \bigcup_{s\in S} B(s)). \tag{3.3.21}$$

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 3.2.10, 3.2.12 and 3.2.13 after substituting δt for 1. Likewise, $G_i^{(t)}(.)$ denotes the t-step transition probability of $\{X'_t\}_{t\in T}$ which is purely generated from $\{G_i^{(\delta t)}(.)\}_{i\in S}$.

Lemma 3.3.14. *Let $\mathscr{B}[\bigcup_{s\in S} B(s)] = \{A \cap \bigcup_{s\in S} B(s) : A \in {}^*\mathscr{B}[X]\}$. Then for any $x \in \bigcup_{s\in S} B(s)$ we have $(\bigcup_{s\in S} B(s), \mathscr{B}[\bigcup_{s\in S} B(s)], g'(x, \delta t, .))$ is an internal probability space.*

Lemma 3.3.15. *For any $i \in S$ and any $t \in T$, we know that $G_i^{(t)}(.)$ is an internal probability measure on $(S, \mathscr{I}(S))$.*

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

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 $\varepsilon_0 \frac{t}{\delta t} \approx 0$ for all $t \in T$). This result is similar to Theorem 3.2.14 but is more complicated.

Theorem 3.3.16. *Suppose (VD) and (SF) hold. Then for any $t \in T$, any $x \in S$ and any near-standard $A \in {}^*\mathscr{B}[X]$, we have*

$$|{}^*g(x,t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(t)}(A\cap S)| \leq \varepsilon_0 + 5\varepsilon_0 \frac{t-\delta t}{\delta t}. \tag{3.3.22}$$

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 $A \in {}^*\mathscr{B}[X]$.

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 $A \in {}^*\mathscr{B}[X]$ and any $x \in S$. By definition, we have:

$$G_x^{(\delta t)}(A\cap S) = g'(x,\delta t,\bigcup_{s'\in A\cap S} B(s')) \tag{3.3.23}$$

$$= {}^*g(x,\delta t,\bigcup_{s'\in A\cap S} B(s')) + \delta_x(\bigcup_{s'\in A\cap S} B(s')){}^*g(x,\delta t,{}^*X \setminus \bigcup_{s\in S} B(s)). \tag{3.3.24}$$

For any $x \in \bigcup_{s'\in A\cap S} B(s')$, by Theorem 3.3.4 and the fact that $\bigcup_{s'\in A\cap S} B(s')$ is near-standard, we have ${}^*g(x,\delta t,{}^*X \setminus \bigcup_{s\in S} B(s)) < \varepsilon_0$ since ${}^*d(x,{}^*X \setminus \bigcup_{s\in S} B(s)) > r_0$. Thus we have $|{}^*g(x,\delta t,\bigcup_{s'\in A\cap S} B(s')) - G_x^{(\delta t)}(A\cap S)| < \varepsilon_0$.

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 $A \in {}^*\mathscr{B}[X]$ 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 3.3.8, we have ${}^*f_x^{(t)}({}^*X \setminus \bigcup_{s\in S} B(s),\bigcup_{s'\in A\cap S} B(s')) < 2\varepsilon_0$. Thus we conclude that:

$$|{}^*g(x,t+\delta t,\bigcup_{s'\in A\cap S}B(s'))-\sum_{s\in S}{}^*g(x,\delta t,B(s)){}^*f_x^{(t)}(B(s),\bigcup_{s'\in A\cap S}B(s'))|<2\varepsilon_0. \tag{3.3.25}$$

By the construction of our hyperfinite representation S and Lemma 3.3.12, we know that for any $s\in S$ we have $|{}^*g(s,t,\bigcup_{s'\in A\cap S}B(s'))-{}^*f_x^{(t)}(B(s),\bigcup_{s'\in A\cap S}B(s'))|<\varepsilon_0$. By the transfer of Lemma 2.1.20, we have that:

$$|\sum_{s\in S}{}^*g(x,\delta t,B(s)){}^*f_x^{(t)}(B(s),\bigcup_{s'\in A\cap S}B(s'))-\sum_{s\in S}{}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S}B(s'))|<\varepsilon_0. \tag{3.3.26}$$

Let us now consider formulas

$$\sum_{s\in S}{}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S}B(s')) \text{ and } \sum_{s\in S}g'(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S}B(s')). \tag{3.3.27}$$

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:

$$|\sum_{s\in S}{}^*g(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S}B(s'))-\sum_{s\in S}g'(x,\delta t,B(s)){}^*g(s,t,\bigcup_{s'\in A\cap S}B(s'))| \tag{3.3.28}$$

$$=|{}^*g(x,\delta t,B(s_0))-g'(x,\delta t,B(s_0))|{}^*g(s_0,t,\bigcup_{s'\in A\cap S}B(s')). \tag{3.3.29}$$

Recall the properties of r_1 constructed after Theorem 3.3.4. If ${}^*d(s_0,y)>r_1$ for all near-standard $y\in \mathrm{NS}({}^*X)$, by Theorem 3.3.4, we have ${}^*g(s_0,t,\bigcup_{s'\in A\cap S}B(s'))<\varepsilon_0$. This implies that

$$|{}^*g(s_0,\delta t,B(s))-g'(s_0,\delta t,B(s))|{}^*g(s_0,t,\bigcup_{s'\in A\cap S}B(s'))<\varepsilon_0. \tag{3.3.30}$$

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 3.3.7, we know that ${}^*g(s_0,\delta t,{}^*X\setminus\bigcup_{s\in S}B(s))<\varepsilon_0$. As $x\in B(s_0)$, by Theorem 3.3.11, we know that

$$|{}^*g(x,\delta t,B(s_0))-g'(x,\delta t,B(s_0))|=|{}^*g(x,\delta t,{}^*X\setminus\bigcup_{s\in S}B(s))|<2\varepsilon_0. \tag{3.3.31}$$

To conclude we have:

$$|\sum_{s\in S} {}^*g(x,\delta t,B(s))^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))-\sum_{s\in S} g'(x,\delta t,B(s))^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))|<2\varepsilon_0. \tag{3.3.32}$$

Finally by induction hypothesis and the transfer of Lemma 2.1.20 we know that:

$$|\sum_{s\in S} g'(x,\delta t,B(s))^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))-G_x^{(t+\delta t)}(A\cap S)| \tag{3.3.33}$$

$$=|\sum_{s\in S} g'(x,\delta t,B(s))^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))-\sum_{s\in S} g'(x,\delta t,B(s))G_s^{(t)}(A\cap S)| \tag{3.3.34}$$

$$\leq |{}^*g(s,t,\bigcup_{s'\in A\cap S} B(s'))-G_s^{(t)}(A\cap S)|\leq \varepsilon_0+5\varepsilon_0\frac{t-\delta t}{\delta t}. \tag{3.3.35}$$

Thus by Eq. (3.3.25),Eq. (3.3.26),Eq. (3.3.32) and Eq. (3.3.35) we conclude that

$$|{}^*g(x,t+\delta t,\bigcup_{s'\in A\cap S} B(s'))-G_x^{(t+\delta t)}(A\cap S)| \tag{3.3.36}$$

$$\leq \varepsilon_0+4\varepsilon_0\frac{t-\delta t}{\delta t}+5\varepsilon_0=\varepsilon_0+5\varepsilon_0\frac{t}{\delta t}. \tag{3.3.37}$$

As all the parameters in this statement are internal, by internal induction principle, we have shown the statement. As $\varepsilon_0\frac{t}{\delta t}\approx 0$ 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 $A\in {}^*\mathscr{B}[X]$. □

As the state space X is σ-compact, by Lemma 2.3.5 and Theorem 2.3.9, we know that $\mathsf{st}^{-1}(A)$ is universally Loeb measurable for $A\in\mathscr{B}[X]$. We now extend Theorem 3.3.16 to establish the relationship between $\overline{{}^*g}$ and $\overline{G}$.

Theorem 3.3.17. *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 (VD) and (SF), for any $E\in\mathscr{B}[X]$ 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. (2.4.19)), for $t > 0$ we have

$$\overline{^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{^*g}(x,t,\bigcup_{s\in A_i} B(s)) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\}. \tag{3.3.38}$$

As the distance between x and s_x is less than δ_0, by Theorem 3.3.11 we know that

$$|^*g(x,t,\bigcup_{s\in A_i} B(s)) - {^*g}(s_x,t,\bigcup_{s\in A_i} B(s))| < \varepsilon_0. \tag{3.3.39}$$

By Theorem 3.3.16, 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

$$\overline{^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{G}_{s_x}(A_i) : A_i \subset \mathsf{st}^{-1}(E)\cap S, A_i \in \mathscr{I}(S)\} = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E)\cap S) \tag{3.3.40}$$

finishing the proof. □

One of the desired properties for a hyperfinite Markov chain is strong regularity. Recall from Definition 3.1.6 that a hyperfinite Markov chain is strong regular if for any $A \in \mathscr{I}(S)$, 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 3.3.18. *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 $A \in {^*\mathscr{B}[X]}$ 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 $\varepsilon \in \mathbb{R}$. 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 $(\forall y \in X)(|y - x_0| < \delta \implies ((\forall A \in \mathscr{B}[X])|g(y,t',A) - g(x_0,t',A)| < \varepsilon))$. By the transfer principle and the fact that $x_1 \approx x_2 \approx x_0$ we know that

$$(\forall A \in {^*\mathscr{B}[X]})(|^*g(x_1,t',A) - {^*g}(x_2,t',A)| < \varepsilon). \tag{3.3.41}$$

As $t' \leq t_0$, by Eq. (3.3.19), we know that $|^*g(x_1,t_0,A) - {^*g}(x_2,t_0,A)| < \varepsilon$ for all $A \in {^*\mathscr{B}[X]}$. 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 $A \in {^*\mathscr{B}[X]}$. □

An immediate consequence of this lemma is the following:

Lemma 3.3.19. *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 3.3.20. *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 $P^{(t)}_{x_0}(.)$ is Radon, we can find K compact, U open with $K \subset E \subset U$ and $P^{(t)}_{x_0}(U) - P^{(t)}_{x_0}(K) < \varepsilon/2$. By the transfer principle, we know that ${}^*g(x_0,t,{}^*U) - {}^*g(x_0,t,{}^*K) < \varepsilon/2$. By Lemma 3.3.18, 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 ${}^*g(x,t,{}^*U) - {}^*g(x,t,{}^*K) < \varepsilon$. 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 $|{}^*g(x,t,{}^*E) - \overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))| < \varepsilon$. This is true for any ε and hence ${}^*g(x,t,{}^*E) \approx \overline{{}^*g}(x,t,\mathsf{st}^{-1}(E))$. $\square$

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 3.3.21. *Suppose (VD) 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 $A \in \mathscr{I}(S)$ 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 3.3.17, we know that $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E) \cap S)$. Fix any $\varepsilon \in \mathbb{R}^+$ and any $A \in \mathscr{I}(S)$, we now consider $G^{t'}_{s_1}(A)$ and $G^{t'}_{s_2}(A)$. By Lemma 3.3.20, we can find a near-standard $A_i \in \mathscr{I}(S)$ such that $|G^{(t')}_{s_1}(A) - G^{(t')}_{s_1}(A_i)| < \frac{\varepsilon}{3}$ and $|G^{(t')}_{s_2}(A) - G^{(t')}_{s_2}(A_i)| < \frac{\varepsilon}{3}$. As A_i is near-standard, by Theorem 3.3.16, 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 3.3.18, 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 $|G^{(t')}_{s_1}(A) - G^{(t')}_{s_2}(A)| < \varepsilon$. 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 $\parallel G^{(t')}_{s_1}(.) - G^{(t')}_{s_1}(.) \parallel \approx 0$ where $\parallel G^{(t')}_{s_1}(.) - G^{(t')}_{s_1}(.) \parallel$ denotes the total variation distance between $G^{(t')}_{s_1}$ and $G^{(t')}_{s_1}$. By Lemma 3.1.25, we know that $\parallel G^{(t)}_{s_1}(.) - G^{(t)}_{s_1}(.) \parallel \approx 0$ hence finishes the proof. $\square$

We are now able to establish to following theorem which is an immediate consequence of Theorem 3.3.21.

Lemma 3.3.22. *Suppose (VD) 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 $\overline{G}_{s_1}^{(t)}(A) = \overline{G}_{s_2}^{(t)}(A)$.*

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 3.3.23. *Suppose (VD) and (SF) hold. For any $s \in \mathrm{NS}(S)$, any non-negative $t \in \mathbb{Q}$ and any $E \in \mathscr{B}[X]$, we have $P^{(t)}_{\mathsf{st}(s)}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. We first prove the theorem when $t = 0$. Fix any $s \in \mathrm{NS}(S)$ and any $E \in \mathscr{B}[X]$. 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 $A \in \mathscr{I}(S)$, 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 $E \in \mathscr{B}[X]$. By Lemmas 3.3.18 and 3.3.20 and Theorem 3.3.17, we know that

$$g(\mathsf{st}(s),t,E) = {}^*g(\mathsf{st}(s),t,{}^*E) \approx {}^*g(s,t,{}^*E) \approx \overline{{}^*g}(s,t,\mathsf{st}^{-1}(E)) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S). \tag{3.3.42}$$

Thus we have for any $s \in \mathrm{NS}(S)$, any non-zero $t \in \mathbb{Q}^+$ and any $E \in \mathscr{B}[X]$: $P^{(t))}_{\mathsf{st}(s)}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$. ∎

It is desirable to extend Theorem 3.3.23 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 3.3.24. 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 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 3.3.25. *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 $\varepsilon \in \mathbb{R}^+$. 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

$$(\forall t \in \mathbb{R}^+)((|t - t_0| < \delta) \implies (|g(x_0, t, U) - g(x_0, t_0, U)| < \varepsilon)). \tag{3.3.43}$$

By the transfer principle, we know that

$$(\forall t \in {}^*\mathbb{R}^+)((|t - t_0| < \delta) \implies (|{}^*g(x_0, t, {}^*U) - {}^*g(x_0, t_0, {}^*U)| < \varepsilon)). \tag{3.3.44}$$

Since ε is arbitrary and $\mathsf{st}(t_1) = \mathsf{st}(t_2) = t_0$, we have

$${}^*g(x_0, t_1, {}^*U) \approx {}^*g(x_0, t_0, {}^*U) \approx {}^*g(x_0, t_2, {}^*U). \tag{3.3.45}$$

By Lemma 3.3.18, we then have

$${}^*g(x_1, t_1, {}^*U) \approx {}^*g(x_0, t_1, {}^*U) \approx {}^*g(x_0, t_2, {}^*U) \approx {}^*g(x_2, t_2, {}^*U), \tag{3.3.46}$$

completing the proof. □

The next lemma establishes the relation between U and $\mathsf{st}^{-1}(U)$.

Lemma 3.3.26. *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 ${}^*g(x, t, {}^*U) - \overline{{}^*g}(x, t, \mathsf{st}^{-1}(U)) < \varepsilon$ for every

$\varepsilon \in \mathbb{R}^+$. Fix some $\varepsilon_1 \in \mathbb{R}^+$. By Lemma 3.3.25, we know that

$$^*g(x,t,{}^*U) \approx {}^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*U). \tag{3.3.47}$$

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

$$^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*U) - {}^*g(\mathsf{st}(x),\mathsf{st}(t),{}^*(\bigcup_{n\leq N})U_n) < \varepsilon_1. \tag{3.3.48}$$

By Lemma 3.3.25 again, we know that ${}^*g(x,t,{}^*U) - {}^*g(x,t,{}^*(\bigcup_{n\leq N})U_n) < \varepsilon_1$. 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 ${}^*g(x,t,{}^*U) - \overline{{}^*g}(x,t,\mathsf{st}^{-1}(U)) < \varepsilon_1$. As the choice of ε_1 is arbitrary, we have the desired result. $\square$

Before we extend Theorem 3.3.23 to all non-negative $t \in \mathbb{R}$, we introduce the following concept.

Definition 3.3.27. A class $\mathscr{C}$ 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 3.3.28 ([22, Lemma 1.17])**.** *Let μ and ν be bounded measures on some measurable space $(\Omega,\mathscr{A})$, and let $\mathscr{C}$ be a π-system in Ω such that $\Omega \in \mathscr{C}$ and $\sigma(\mathscr{C}) = \mathscr{A}$ where $\sigma(\mathscr{C})$ denote the σ-algebra generated by $\mathscr{C}$. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ for all $A \in \mathscr{C}$.*

Lemma 3.3.28 allows us to obtain slightly stronger results than Lemmas 3.3.25 and 3.3.26.

Lemma 3.3.29. *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 $E \in \mathscr{B}[X]$, 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 $A \in \mathscr{B}[X]$. It is easy to see that both μ_1 and μ_2 are probability measures on X. By Lemma 3.3.25, 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 3.3.28, we have the desired result. $\square$

By using essentially the same argument, we have

Lemma 3.3.30. *Suppose (SF) and (OC) hold. For any $E \in \mathscr{B}[X]$, 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 3.3.23 to all non-negative $t \in \mathbb{R}$.

Theorem 3.3.31. *Suppose (VD), (SF) and (OC) hold. For any $s \in \mathrm{NS}(S)$, any non-infinitesimal $t \in \mathrm{NS}(T)$ and any $E \in \mathscr{B}[X]$, we have $P_{\mathsf{st}(s)}^{(\mathsf{st}(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$.*

Proof. Fix any $s \in \mathrm{NS}(S)$, any non-infinitesimal $t \in \mathrm{NS}(T)$ and any $E \in \mathscr{B}[X]$. By Lemmas 3.3.29 and 3.3.30, we know that

$$g(\mathsf{st}(s), \mathsf{st}(t), E) = {}^*g(\mathsf{st}(s), \mathsf{st}(t), {}^*E) \approx {}^*g(s, t, {}^*E) \approx \overline{{}^*g}(s, t, \mathsf{st}^{-1}(E)). \tag{3.3.49}$$

By Theorem 3.3.17, we know that $\overline{{}^*g}(s, t, \mathsf{st}^{-1}(E)) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$. Thus we know that $g(\mathsf{st}(s), \mathsf{st}(t), E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$, 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 3.3.31, we can show that $g(\mathsf{st}(s), \mathsf{st}(t), U) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(U) \cap S)$ 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 3.3.32 ([13, 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 3.3.33 ([36])**.** *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 $E \in \mathscr{B}[X]$ 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 3.3.34. *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 3.3.25. $\square$

This lemma, just like Lemma 3.3.25, is stronger than Lemma 3.3.18 since t_1 and t_2 need not be standard positive real numbers. We now generalize Lemma 3.3.20 to all $t \in \mathrm{NS}({}^*\mathbb{R})$. Before proving it, we first recall the following theorem.

Theorem 3.3.35 (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*

$$\lim_{n\to\infty} \mu_n(X) = \mu(X). \tag{3.3.50}$$

exists for every $X \in \Sigma$, then μ is countably additive.

An immediate consequence of Theorem 3.3.35 is that the limit of probability measures remain a probability measure. The following lemma generalizes Lemma 3.3.20 to all $t \in \mathrm{NS}({}^*\mathbb{R})$.

Lemma 3.3.36. *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 ${}^*g(x,t,{}^*E) \approx \overline{{}^*g}(x,t,\mathrm{st}^{-1}(E))$ for all $E \in \mathscr{B}[X]$. Moreover, $\overline{{}^*g}(x,t,\mathrm{st}^{-1}(X)) = 1$ 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 $E \in \mathscr{B}[X]$. Let $x_0 = \mathrm{st}(x)$ and $t_0 = \mathrm{st}(t)$. We first show the result for $t < t_0$. For any $B \in \mathscr{B}[X]$, 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 $(X,\mathscr{B}[X])$. Since X is a Polish space, h is a Radon measure. By Lemma 2.4.8, we know that $\overline{{}^*h}(x_0,t_0,\mathrm{st}^{-1}(X)) = 1$. As $t \approx t_0$, we know that ${}^*g(x_0,t,{}^*B) \approx {}^*h(x_0,t_0,{}^*B)$ for all $B \in \mathscr{B}[X]$. Pick some $\varepsilon \in \mathbb{R}^+$ and choose K compact, U open with $K \subset E \subset U$ and $h(x_0,t_0,U) - h(x_0,t_0,K) < \frac{\varepsilon}{2}$. We have

$$|\overline{{}^*g}(x_0,t,\mathrm{st}^{-1}(E)) - \overline{{}^*h}(x_0,t_0,\mathrm{st}^{-1}(E))| \tag{3.3.51}$$

$$\lessapprox |\overline{{}^*g}(x_0,t,\mathrm{st}^{-1}(E)) - \overline{{}^*g}(x_0,t,{}^*K)| + |\overline{{}^*h}(x_0,t_0,{}^*K) - \overline{{}^*h}(x_0,t_0,\mathrm{st}^{-1}(E))| \lessapprox \varepsilon \tag{3.3.52}$$

As ε is arbitrary, we have $\overline{{}^*g}(x_0,t,\mathrm{st}^{-1}(E)) = \overline{{}^*h}(x_0,t_0,\mathrm{st}^{-1}(E))$. Hence we have $\overline{{}^*g}(x_0,t,\mathrm{st}^{-1}(E)) = {}^*g(x_0,t,{}^*E)$. By Lemma 3.3.18, we know that ${}^*g(x_0,t,D) \approx {}^*g(x,t,D)$ for all $D \in {}^*\mathscr{B}[X]$. Thus, we have $\overline{{}^*g}(x_0,t,\mathrm{st}^{-1}(E)) = \overline{{}^*g}(x,t,\mathrm{st}^{-1}(E))$ and ${}^*g(x_0,t,{}^*E) \approx {}^*g(x,t,{}^*E)$. Hence we have $\overline{{}^*g}(x,t,\mathrm{st}^{-1}(E)) = {}^*g(x,t,{}^*E)$.

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 $B \in \mathscr{B}[X]$.

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 $A \in {}^*\mathscr{B}[X]$ such that

$$(A \cap \mathsf{st}^{-1}(X) = \emptyset) \wedge ({}^*g(x_0,t_0,A) > \frac{1}{n}). \tag{3.3.53}$$

Pick some positive $t_1 \in \mathbb{R}$.

Claim 3.3.37. ${}^*f_{x_0}^{(t_0,t_1)}(A,{}^*K) \approx 0$ *for all compact* $K \subset X$.

Proof. Pick some compact subset K and some positive $\varepsilon \in \mathbb{R}$. By condition (2) of (VD), there exists positive $r \in \mathbb{R}$ such that

$$(\forall x \in X)(d(x,K) > r \implies g(x,t_1,K) < \varepsilon). \tag{3.3.54}$$

By the transfer principle, we know that ${}^*g(x,t_1,{}^*K) \approx 0$ for all $x \in A$. By Lemma 3.2.5, we have ${}^*f_{x_0}^{(t_0,t_1)}(A,{}^*K) \approx 0$. □

Fix some compact $K \subset X$. Note that

$${}^*g(x_0,t_0+t_1,K) = {}^*g(x_0,t_0,A){}^*f_{x_0}^{(t_0,t_1)}(A,{}^*K) + {}^*g(x_0,t_0,{}^*X \setminus A){}^*f_{x_0}^{(t_0,t_1)}({}^*X \setminus A,{}^*K). \tag{3.3.55}$$

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 3.3.38. *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 3.3.17 and Lemma 3.3.36, we know that

$$\overline{G}_s^{(t)}(\mathsf{st}^{-1}(X) \cap S) = {}^*g(s,t,\mathsf{st}^{-1}(X)) = 1. \tag{3.3.56}$$

□

Assuming (WC) instead of (OC), we have the following result which is similar to Theorem 3.3.31.

Theorem 3.3.39. *Suppose (VD), (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 $E \in \mathscr{B}[X]$, 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 3.3.33. The proof is similar to the proof of Theorem 3.3.31 hence omitted. □

Chapter 4

Convergence Results for Standard Markov Processes

In the previous chapter, for a continuous-time general state space Markov process $\{X_t\}_{t\geq 0}$ satisfying certain regularity conditions, we have constructed a hyperfinite Markov process $\{X'_t\}_{t\in T}$ such that the internal transition probabilities of $\{X'_t\}_{t\in T}$ differs from the transition probabilities of $\{X_t\}_{t\geq 0}$ only by infinitesimal. Such $\{X'_t\}_{t\in T}$ is called a hyperfinite representation of $\{X_t\}_{t\geq 0}$. In this chapter, we will establish some convergence results for the standard Markov process $\{X_t\}_{t\geq 0}$ using convergence result on $\{X'_t\}_{t\in T}$ in Section 3.1.

In Section 4.1, 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 4.1.6 and Lemmas 4.1.8 and 4.1.15). By Theorem 3.1.26, we know that $\{X'_t\}_{t\in T}$ is ergodic. The ergodicity of $\{X_t\}_{t\in T}$ (Theorem 4.1.16) follows from Theorem 3.1.26. It will be shown in Example 5.3.8 that the Markov chain ergodic theorem established in this dissertation is incomparable to the existing result (Theorem 5.3.7) in the literature.

One of the major assumptions on $\{X_t\}_{t\geq 0}$ is the strong Feller assumption 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 assuming 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 4.2, 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 open whether the Markov chain ergodic theorem is true when $\{X_t\}_{t\geq 0}$ only satisfies the Feller condition.

4.1 Markov Chain Ergodic Theorem

In the last section, we established the relation 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 4.1.1. *For any $t\in\mathbb{R}^+$, any finite partition of X with Borel sets $A_1,....,A_n,B$ of X and any $A\in\mathscr{B}[X]$ such that:*

1. *for each $A_i\in\{A_1,....,A_n\}$ there exists an $\varepsilon_i\in\mathbb{R}^+$ such that for any $x,y\in A_i$ we have $|P^{(t)}_x(A)-P^{(t)}_y(A)|<\varepsilon_i$.*
2. *there exists an $\varepsilon\in\mathbb{R}^+$ such that $\pi(B)<\varepsilon$.*

We have $|\pi(A)-\sum_{i\leq n}\pi(A_i)P^{(t)}_{x_i}(A)|\leq\sum_{i\leq n}\pi(A_i)\varepsilon_i+\varepsilon$ 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 $A\in\mathscr{B}[X]$ and any $x_i\in A_i$. We then have:

$$|\pi(A)-\sum_{i\leq n}\pi(A_i)P^{(t)}_{x_i}(A)| \tag{4.1.1}$$

$$=|\int_X P^{(t)}_x(A)\pi(dx)-\sum_{i\leq n}(\int_{A_i}\pi(dx))P^{(t)}_{x_i}(A)| \tag{4.1.2}$$

$$=|\sum_{i\leq n}\int_{A_i}P^{(t)}_x(A)\pi(dx)+\int_B P^{(t)}_x(A)\pi(dx)-\sum_{i\leq n}\int_{A_i}P^{(t)}_{x_i}(A)\pi(dx)| \tag{4.1.3}$$

$$\leq|\sum_{i\leq n}(\int_{A_i}(P^{(t)}_x(A)-P^{(t)}_{x_i}(A))\pi(dx))|+\varepsilon \tag{4.1.4}$$

$$\leq\sum_{i\leq n}(\int_{A_i}\varepsilon_i\pi(dx))+\varepsilon \tag{4.1.5}$$

$$=\sum_{i\leq n}\pi(A_i)\varepsilon_i+\varepsilon. \tag{4.1.6}$$

□

Write $P_x^{(t)}(A)$ as $g(x,t,A)$ and then apply the transfer principle, we have the following lemma:

Lemma 4.1.2. *For any $t \in {}^*\mathbb{R}^+$, for any hyperfinite partition of ${}^*X$ with *Borel sets $A_1,....,A_N,B$ of ${}^*X$ and any $A \in {}^*\mathscr{B}[X]$ such that:*

1. *for each $A_i \in \{A_1,....,A_N\}$ there exists an $\varepsilon_i \in {}^*\mathbb{R}^+$ such that for any $x,y \in A_i$ $|{}^*g(x,t,A) - {}^*g(x,t,A)| < \varepsilon_i$.*
2. *there exists an $\varepsilon \in {}^*\mathbb{R}^+$ such that $\pi(B) < \varepsilon$.*

We have

$$|{}^*\pi(A) - \sum_{i \leq N} {}^*\pi(A_i){}^*g(x_i,t,A)| \leq \sum_{i \leq N} {}^*\pi(A_i)\varepsilon_i + \varepsilon. \tag{4.1.7}$$

for any $x_i \in A_i$

Recall the definition of weakly stationary distribution:

Definition 4.1.3. An internal distribution π' on $(S,\mathscr{I}(S))$ 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 $A \in \mathscr{I}(S)$ 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 4.1.4. Define an internal probability measure π' on $(S,\mathscr{I}(S))$ 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 4.1.4.

Lemma 4.1.5. *π' is an internal probability measure on $(S,\mathscr{I}(S))$. Moreover, for any $A \in \mathscr{B}[X]$, we have $\pi(A) = \overline{\pi'}(\mathsf{st}^{-1}(A) \cap S)$.*

Proof. Clearly π' is an internal measure on $(S,\mathscr{I}(S))$. The second part of the lemma follows directly from Theorem 2.4.11. □

We now show that π' is a weakly stationary distribution for $\{X'_t\}$.

Theorem 4.1.6. *Suppose (VD), (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 $\mathscr{F} = \{B(s_1),....,B(s_N),{}^*X \setminus \bigcup_{s\in S} B(s)\}$ of ${}^*X$ where $S = \{s_1,s_2,...,s_N\}$ is the state space of $\{X'_t\}$. Note that every member of $\mathscr{F}$ is an member of ${}^*\mathscr{B}[X]$. By Theorem 3.3.11 and Eq. (3.3.19), we know that

$$(\forall i \le N)(\forall x,y \in B(s_i))(\forall C \in {}^*\mathscr{B}[X])(|{}^*g(x,t,C) - {}^*g(y,t,C)| < \varepsilon_0). \tag{4.1.8}$$

Let $B = \bigcup_{s\in A} B(s)$ then $B \in {}^*\mathscr{B}[X]$ 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 ${}^*\pi({}^*X \setminus \bigcup_{s\in S} B(s)) = \varepsilon_1$.

By Lemma 4.1.2 , we have

$$|{}^*\pi(B) - \sum_{i\le N} {}^*\pi(B(s_i)){}^*g(s_i,t,B)| \le \sum_{i\le N} {}^*\pi(B(s_i))\varepsilon_0 + \varepsilon_1 \le \varepsilon_0 + \varepsilon_1 \approx 0. \tag{4.1.9}$$

By Definition 4.1.4, 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

$$|\pi'(A) - \sum_{i\le N} \pi'(s_i){}^*g(s_i,t,B)| \approx 0. \tag{4.1.10}$$

Fix positive $\varepsilon \in \mathbb{R}$. As $\overline{\pi'}$ concentrates on $\mathrm{NS}(S)$, there is a near-standard internal set C with $\pi'(C) > 1-\varepsilon$. Thus we have

$$|\sum_{s\in S} \pi'(\{s\}){}^*g(s_i,t,B) - \sum_{s\in C} \pi'(\{s\}){}^*g(s_i,t,B)| < \varepsilon \tag{4.1.11}$$

Claim 4.1.7. *Suppose (VD), (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 3.3.38, there exist near-standard $A_i \in \mathscr{I}(S)$ with $A_i \subset A$ such that $G_s^{(t)}(A) - G_s^{(t)}(A_i) < \frac{1}{n_0}$. By Lemma 3.3.36, there exist near-standard $C_i \in {}^*\mathscr{B}[X]$ 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 $B_i \in \mathscr{I}(S)$ 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 3.3.16, we have

$$^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s')) \approx G_s^{(t)}(A_i\cup B_i) \tag{4.1.12}$$

Thus we have

$$|^*g(s,t,B) - G_s^{(t)}(A)| \tag{4.1.13}$$

$$\approx |^*g(s,t,B) - {}^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s')) + G_s^{(t)}(A_i\cup B_i) - G_s^{(t)}(A)| \tag{4.1.14}$$

$$\leq |^*g(s,t,B) - {}^*g(s,t,\bigcup_{s'\in A_i\cup B_i} B(s'))| + |G_s^{(t)}(A_i\cup B_i) - G_s^{(t)}(A)| < \frac{2}{n_0} \tag{4.1.15}$$

As the choice of n_0 is arbitrary, we have the desired result. □

As C is near-standard, by Lemma 2.1.20, we have

$$|\sum_{s\in C} \pi'(\{s\})^*g(s_i,t,B) - \sum_{s\in C} \pi'(\{s\})G_s^{(t)}(A)| \approx 0. \tag{4.1.16}$$

By the construction of C again, we have

$$|\sum_{s\in C} \pi'(\{s\})G_s^{(t)}(A) - \sum_{s\in S} \pi'(\{s\})G_s^{(t)}(A)| < \varepsilon. \tag{4.1.17}$$

By Eqs. (4.1.10), (4.1.11), (4.1.16) and (4.1.17), we have

$$|\pi'(A) - \sum_{s\in S} \pi'(\{s\})G_s^{(t)}(A)| < 2\varepsilon. \tag{4.1.18}$$

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 $\mathscr{D} = \{t \in T : (\forall A \in \mathscr{I}(S))(|\pi'(A) - \sum_{s\in S} \pi'(\{s\})G_s^{(t)}(A)| < \frac{1}{t})\}$. This is an internal set and contains all $t \in \mathrm{NS}(T)$. Suppose there is no infinite t_0 such that $\mathscr{D}$ contains all the infinite t no greater than t_0. This implies $T \setminus \mathscr{D}$ contains arbitrarily small infinite element hence, by underspill, $T \setminus \mathscr{D}$ contains some $t_0 \in \mathrm{NS}(T)$. This contradicts with the fact that $\mathscr{D}$ 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 4.1.8. *Suppose (VD) 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 4.1.6, $(\pi\times\pi)'$ 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

$$(\pi\times\pi)'(\{(s_1,s_2)\}) = \frac{{}^*(\pi\times\pi)(B(s_1)\times B(s_2))}{{}^*(\pi\times\pi)(\bigcup_{s\in S}B(s)\times\bigcup_{s\in S}B(s))} \tag{4.1.19}$$

$$= \frac{{}^*\pi(B(s_1))}{{}^*\pi(\bigcup_{s\in S}B(s))}\cdot\frac{{}^*\pi(B(s_2))}{{}^*\pi(\bigcup_{s\in S}B(s))} \tag{4.1.20}$$

$$= \pi'(s_1)\pi'(s_2). \tag{4.1.21}$$

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 3.1.19, 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 3.1.14, 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 4.1.9. 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 $P_x^{(t)}(B) > 0$.

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 4.1.10. 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 4.1.11. *Suppose (VD), (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 $\overline{G}_s^{(t)}(B)>0$.*

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 4.1.12. $\mathsf{st}^{-1}(A)\cap S\subset B$.

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 $P_x^{(t)}(A)>0$. Pick $t'\in T$ such that $t'\approx t$ and $t'\geq t$. By Lemma 3.3.34, Lemma 3.3.36 and Theorem 3.3.17, we know that

$$P_x^{(t)}(A)=g(x,t,A)={}^*g(x,t,{}^*A)\approx {}^*g(s,t',{}^*A)\approx \overline{{}^*g}(s,t',\mathsf{st}^{-1}(A))=\overline{G}_s^{(t')}(\mathsf{st}^{-1}(A)\cap S). \tag{4.1.22}$$

Then we have $\mathsf{st}((G_s^{(t)}(B)))>0$. □

Let $\{X'_t\}_{t\in T}$ and $\{Y'_t\}_{t\in T}$ be two i.i.d hyperfinite Markov chains on $(S,\mathscr{I}(S))$ both with internal transition probability $\{G_i^{(\delta t)}j)\}_{i,j\in S}$. Let $\{Z'_t\}_{t\in T}$ be the product hyperfinite Markov chain live on $(S\times S,\mathscr{I}(S\times S))$ 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

$$F_{(i,j)}^{(\delta t)}(\{(a,b)\})=G_i^{(\delta t)}(\{a\})\times G_j^{(\delta t)}(\{b\}). \tag{4.1.23}$$

where $(F_{(i,j)}^{(\delta t)}(\{(a,b)\})$ 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 4.1.13. 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 4.1.14. *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 $P_x^{(t)}(A) > 0$ 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 4.1.15. *Suppose (VD), (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 4.1.11, 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 3.3.23, we have

$$Q^{(t)}_{(x,y)}(A_1 \times A_2) = P_x^{(t)}(A_1) \times P_y^{(t)}(A_2) = \overline{G}_i^{(t)}(\mathsf{st}^{-1}(A_1) \cap S) \times \overline{G}_j^{(t)}(\mathsf{st}^{-1}(A_2) \cap S). \tag{4.1.24}$$

By Lemma 3.1.10 and the construction of Loeb measure, we know that

$$\overline{G}_i^{(t)}(\mathsf{st}^{-1}(A_1) \cap S) \times \overline{G}_j^{(t)}(\mathsf{st}^{-1}(A_2) \cap S) = \overline{F}^{(t)}_{(i,j)}(\mathsf{st}^{-1}(A_1 \times A_2)) \cap (S \times S)). \tag{4.1.25}$$

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. □

Now we are at the place to prove the main theorem of this paper.

Theorem 4.1.16. *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 $\{X_t\}_{t\geq 0}$ also satisfies (VD), (SF) and (WC). Then for π-almost surely $x\in X$ we have $\lim_{t\to\infty}\sup_{A\in\mathscr{B}[X]}|P^{(t)}_x(A)-\pi(A)|=0$.*

Proof. 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 4.1.4. By Theorem 4.1.6, 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 4.1.15, we know that $\{X'_t\}_{t\in T}$ is productively near-standard open set irreducible. By Theorem 3.3.21, we know that $\{X'_t\}_{t\in T}$ is strong regular. Thus by Theorems 3.1.19 and 3.1.26, we know that for $\overline{\pi'}$ almost surely $s\in S$ and any $A\in\mathscr{L}(\mathscr{I}(S))$, $\lim_{t\to\infty}\sup_{B\in\mathscr{L}(\mathscr{I}(S))}|\overline{G}^{(t)}_s(B)-\overline{\pi'}(B)|=0$.

Now fix any $A\in\mathscr{B}[X]$. Then by Theorem 2.3.9, we know that $\mathsf{st}^{-1}(A)\in\mathscr{L}(\mathscr{I}(S))$. Consider any $x\in X$ and any $s\in\mathsf{st}^{-1}(\{x\})\cap S$. By Theorem 3.3.23, we know that for any $t\in\mathbb{Q}^+$ we have $P^{(t)}_x(A)=\overline{G}^{(t)}_s(\mathsf{st}^{-1}(A)\cap S)$. By Lemma 4.1.5, we know that $\pi(A)=\overline{\pi'}(\mathsf{st}^{-1}(A)\cap S)$. Suppose that there exists a set $B\in\mathscr{B}[X]$ with $\pi(B)>0$ such that, for any $x\in B$, $P^{(t)}_x(.)$ does not converge to $\pi(.)$ in total variation distance. This means that for any $s\in\mathsf{st}^{-1}(B)\cap S$ we have

$$\sup_{A\in\mathscr{B}[X]}|\overline{G}^{(t)}_s(\mathsf{st}^{-1}(A)\cap S)-\overline{\pi'}(\mathsf{st}^{-1}(A)\cap S)|\nrightarrow 0. \tag{4.1.26}$$

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, $\lim_{t\to\infty}\sup_{B\in\mathscr{L}(\mathscr{I}(S))}|\overline{G}^{(t)}_s(B)-\overline{\pi'}(B)|=0$. Hence we have the desired result. □

4.2 The Feller Condition

In Sections 3.2 and 4.1, 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 4.2.1. (SF) The Markov chain$\{X_t\}_{t\geq 0}$ is said to be *strong Feller* if for any $t>0$ and any $\varepsilon>0$ we have:

$$(\forall x\in X)(\exists\delta>0)((\forall y\in X)(|y-x|<\delta \implies (\forall A\in\mathscr{B}[X])|P_y^{(t)}(A)-P_x^{(t)}(A)|<\varepsilon)). \tag{4.2.1}$$

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 $\varepsilon>0$ we have:

$$(\forall A\in\mathscr{B}[X])(\forall x\in X)(\exists\delta>0)((\forall y\in X)(|y-x|<\delta \implies |P_y^{(t)}(A)-P_x^{(t)}(A)|<\varepsilon)). \tag{4.2.2}$$

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 4.2.2 (suggested by Neal Madras)**.** [30, 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 4.2.3. $\lim_{n\to\infty}P_{\frac{1}{n}}(A)=\mu(A)$ *for all Borel sets A.*

Proof. Let A be an internal with end points a and b.

$$\lim_{n\to\infty}P_{\frac{1}{n}}(A)=\lim_{n\to\infty}\int_a^b\frac{1+\sin(ny)}{2\pi}\mathrm{d}y \tag{4.2.3}$$

$$=\lim_{n\to\infty}(\frac{b-a}{2\pi}-\frac{\cos(nb)-\cos(na)}{2n\pi}) \tag{4.2.4}$$

$$=\frac{b-a}{2\pi}=\mu(A) \tag{4.2.5}$$

By Theorem 3.3.33, we have the desired result. □

Claim 4.2.4. $\sup_{A\in\mathscr{B}[[-\pi,\pi]]}|P_{\frac{1}{n}}(A)-\mu(A)|\geq\frac{1}{\pi}$ *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. □

4.2.1 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 $P_x^{(t)}(A)$ 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 2.4.3, 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 $\varepsilon_0\frac{t}{\delta t}\approx 0$ for all $t\in T$. We can pick r_0 the same way as we did in Section 3.2. Recall (VD) and Theorem 3.3.4 from Section 3.2.

Remark 4.2.5 ((VD)). The Markov chain $\{X_t\}_{t\geq 0}$ is said to be *vanish in distance* if for all $t\geq 0$ and all $K\in\mathscr{K}[X]$ we have:

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

where $\mathscr{K}$ denote the collection of all compact sets of X.

We have the following lemma from the above definition.

Lemma 4.2.6 (Theorem 3.3.4)**.** *Suppose (VD) holds. For every positive $\varepsilon\in{}^*\mathbb{R}$, there exists an open ball centered at some standard point a with radius r such that:*

1. ${}^*g(x,\delta t,{}^*X\setminus\overline{U}(a,r))<\varepsilon$ *for all* $x\in \mathrm{NS}({}^*X)$.
2. ${}^*g(y,t,A)<\varepsilon$ *for all* $y\in{}^*X\setminus\overline{U}(a,r)$, *all near-standard* $A\in{}^*\mathscr{B}[X]$ *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 4.2.6.

Recall the following results from Section 3.2

Lemma 4.2.7 (Lemma 3.3.7)**.** *Suppose (VD) holds. There exists a positive infinite $r_0 > 2r_1$ such that*

$$(\forall y \in \overline{U}(a_0, 2r_1))({}^*g(y, \delta t, {}^*X \setminus \overline{U}(a_0, r_0)) < \varepsilon_0). \tag{4.2.6}$$

Just as in Section 3.2,we fix a_0, r_1 and r_0 for the remainder of this section.

Lemma 4.2.8 (Lemma 3.3.8)**.** *Suppose (VD) holds. For any $x \in X$, any $t \in T$, any near-standard internal set $A \subset {}^*X$ we have ${}^*f_x^{(t)}({}^*X \setminus \overline{U}(a_0, r_0), A) < 2\varepsilon_0$.*

Just as in Section 3.2, 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 4.2.9. *Suppose (VD) 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 $A \in \mathscr{I}(S_1)$ and all $t \in T^+$:*

$$|{}^*g(x_1, t, \bigcup_{s \in A} B_1(s)) - {}^*g(x_2, t, \bigcup_{s \in A} B_1(s))| < \varepsilon_0 \tag{4.2.7}$$

Proof. Fix a $A \in \mathscr{I}(S_1)$. 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

$$|y - x| < \delta_x \implies |{}^*g(x, \delta t, \bigcup_{s \in A} B_1(s)) - {}^*g(y, \delta t, \bigcup_{s \in A} B_1(s))| < \frac{\varepsilon_0}{2}. \tag{4.2.8}$$

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 $\mathscr{F} = \{U(x_i, \frac{\delta_{x_i}}{2}) : i \leq N\}$ where $\{\frac{\delta_{x_i}}{2} : i \leq N\}$ is a hyperfinite set. Let $\delta_A = \min\{\frac{\delta_{x_i}}{2} : i \leq N\}$.

Pick any $x, y \in \overline{U}(a_0, 2r_0)$ with $|x - y| < \delta_A$. We know that $x \in B(x_i, \frac{\delta_{x_i}}{2})$ for some $i \leq N$ and ${}^*d(y, x_i) \leq {}^*d(y, x) + {}^*d(x, x_i) \leq \delta_{x_i}$. Thus both x, y are in some $B(x_i, \delta_{x_i})$. This means that

$$|{}^*g(x, \delta t, \bigcup_{s \in A} B_1(s)) - {}^*g(y, \delta t, \bigcup_{s \in A} B_1(s))| < \varepsilon_0. \tag{4.2.9}$$

Let $\mathscr{M} = \{\delta_A : A \in \mathscr{I}(S)\}$. Note that $\mathscr{M}$ 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. □

Definition 4.2.10. Let S, S' be two hyperfinite representations of ${}^*X$. The hyperfinite representation S' is a refinement of S if for every $A \in \mathscr{I}(S)$ there exists a $A' \in \mathscr{I}(S')$ 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 $A \in \mathscr{I}(S)$, note that the enlargement A' is unique. Fix δ_0 in Lemma 4.2.9 for the remainder of this section. We present the following result.

Lemma 4.2.11. *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

$$M(s) = \{B_H(h) : B_H(h) \cap B_1(s) \neq \emptyset\}. \tag{4.2.10}$$

Note that $M(s)$ is hyperfinite for every $s \in S_1$. Let

$$N(s) = \{B_H(h) \cap B_1(s) : B_H(h) \in M(s)\}. \tag{4.2.11}$$

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. □

For each $s \in S$, we use $B(s)$ to denote the corresponding *Borel set. By the construction in Lemma 4.2.11, 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 4.2.9 and 4.2.11, we have the following result:

Theorem 4.2.12. *Let S_1, S be the same hyperfinite representations as in Lemma 4.2.11. Then for any $s \in S$, any $x_1, x_2 \in B(s)$, any $A \in \mathscr{I}(S_1)$ and any $t \in T^+$ we have*

$$|{}^*g(x_1, t, \bigcup_{s\in A} B_1(s)) - {}^*g(x_2, t, \bigcup_{s\in A} B_1(s))| < \varepsilon_0. \tag{4.2.12}$$

An immediate consequence of this theorem is:

Corollary 4.2.13. *Let S_1, S be the same hyperfinite representations as in Lemma 4.2.11. For for any $s \in S$, any $y \in B(s)$, any $x \in {}^*X$, any $A \in \mathscr{I}(S_1)$ and any $t \in T^+$ we have $|{}^*g(y,t,\bigcup_{s\in A} B_1(s)) - {}^*f_x^{(t)}(B(s), \bigcup_{s\in A} B_1(s))| < \varepsilon_0$.*

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 3.2. 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 $G_i^{(t)}(.)$ to denote the t-step internal transition measure. By Lemmas 3.2.12 and 3.2.13, we know that $G_i^{(t)}(.)$ is an internal probability measure on $(S, \mathscr{I}(S))$ for all $t \in T$.

Similar to Theorem 3.3.16, we have the following theorem. The two proofs are similar to each other.

Theorem 4.2.14. *Suppose (VD) and (WF) hold. For any $t \in T$, any $x \in S$ and any near-standard $A \in \mathscr{I}(S_1)$, we have*

$$|{}^*g(x,t,\bigcup_{s\in A_S} B(s)) - G_x^{(t)}(A_S)| \leq \varepsilon_0 + 5\varepsilon_0 \frac{t-\delta t}{\delta t}. \tag{4.2.13}$$

where A_S is the enlargement of A. In particular, for all $t \in T$, all $x \in S$ and all near-standard $A \in \mathscr{I}(S_1)$ we have

$$|{}^*g(x,t,\bigcup_{s\in A_S} B(s)) - G_x^{(t)}(A_S)| \approx 0 \tag{4.2.14}$$

Proof. : In the proof of Theorem 3.3.16, by (SF), we know that for any $s_0 \in S$ and any $t \in T+$

$$(\forall x_1, x_2 \in B(s_0))(\forall A \in \mathscr{I}(S))(|{}^*g(x_1,t,\bigcup_{s\in A} B(s)) - {}^*g(x_2,t,\bigcup_{s\in A} B(s))| < \varepsilon_0). \tag{4.2.15}$$

Under (WF), by Theorem 4.2.12 and Corollary 4.2.13 and the fact that S is a refinement of S_1, we know that for any $s_0 \in S$ and any $t \in T^+$

$$(\forall x_1, x_2 \in B(s_0))(\forall A \in \mathscr{I}(S_1))(|{}^*g(x_1,t,\bigcup_{s\in A_S} B(s)) - {}^*g(x_2,t,\bigcup_{s\in A_S} B(s))| < \varepsilon_0). \tag{4.2.16}$$

We use this formula to replace the Eq. (4.2.15) in the proof of Theorem 3.3.16. Then the rest of the proof is identical to the proof of Theorem 3.3.16. □

In Section 3.2, 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 3.3.17.

Theorem 4.2.15. *Suppose (VD) 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 $E \in \mathscr{B}[X]$ and any $t \in T$, we have $\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E)\cap S)$.*

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 $\overline{{}^*g}(x,0,\mathsf{st}^{-1}(E)) = \overline{G}^{(0)}_{s_x}(\mathsf{st}^{-1}(E)\cap S)$.

We now prove the case for $t > 0$. Fix some $x \in \bigcup_{s\in S} B(s)$, some $t > 0$ and some $E \in \mathscr{B}[X]$. By the construction in Theorem 2.4.11 and Eq. (2.4.19), we know that for every $t > 0$:

$$\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{{}^*g}(x,t,\bigcup_{s\in A} B_1(s)) : A \subset \mathsf{st}^{-1}(E)\cap S_1, A \in \mathscr{I}(S_1)\} \tag{4.2.17}$$

By Theorem 4.2.12, we have $|{}^*g(x,t,\bigcup_{s\in A} B_1(s)) - {}^*g(s_x,t,\bigcup_{s\in A} B_1(s))| < \varepsilon_0$. By Theorem 4.2.14, 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

$$\overline{{}^*g}(x,t,\mathsf{st}^{-1}(E)) = \sup\{\overline{G}^{(t)}_{s_x}(A_S) : A \subset \mathsf{st}^{-1}(E)\cap S_2, A \in \mathscr{I}(S_1)\}. \tag{4.2.18}$$

Claim 4.2.16.

$$\overline{G}^{(t)}_{s_x}(\mathsf{st}^{-1}(E)\cap S) = \sup\{\overline{G}^{(t)}_{s_x}(A_S) : A \subset \mathsf{st}^{-1}(E)\cap S_2, A \in \mathscr{I}(S_1)\}. \tag{4.2.19}$$

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 $A \in \mathscr{I}(S_1)$ and it is easy to see that $B \subset A_S \subset$

$\mathsf{st}^{-1}(E)\cap S$. Thus we can conclude that

$$\sup\{\overline{G}_{s_x}^{(t)}(A_S) : A \subset \mathsf{st}^{-1}(E)\cap S_2, A \in \mathscr{I}(S_2)\} = \overline{G}_{s_x}^{(t)}(\mathsf{st}^{-1}(E)\cap S). \tag{4.2.20}$$

□

Thus we have the desired result. □

The next lemma establishes a weaker form of local continuity of ${}^*g$.

Lemma 4.2.17. *Suppose (WF) holds. For any two near-standard $x_1 \approx x_2$ from ${}^*X$,any $t \in \mathbb{R}^+$ and any $A \in \mathscr{B}[X]$ 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 $A \in \mathscr{B}[X]$. Pick $\varepsilon \in \mathbb{R}^+$. By (WF), we can pick a $\delta \in \mathbb{R}^+$ such that

$$(\forall y \in X)(|y - x_0| < \delta \implies (|g(y,t,A) - g(x_0,t,A)| < \varepsilon)). \tag{4.2.21}$$

By the transfer principle and the fact that $x_1 \approx x_2 \approx x_0$ we know that

$$(|{}^*g(x_1,t,{}^*A) - {}^*g(x_2,t,{}^*A)| < \varepsilon). \tag{4.2.22}$$

As ε is arbitrary, this completes the proof. □

As Lemma 3.3.20, the next lemma establishes the link between ${}^*E$ and $\mathsf{st}^{-1}(E)$ for every $E \in \mathscr{B}[X]$.

Lemma 4.2.18. *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 4.2.17 and is similar to the proof of Lemma 3.3.20. □

Lemmas 4.2.17 and 4.2.18 allow us to obtain the result in Theorem 3.3.23 under weaker assumptions.

Theorem 4.2.19. *Suppose (VD) and (WF) hold. For any $s \in \mathrm{NS}(S)$, any non-negative $t \in \mathbb{Q}$ and any $E \in \mathscr{B}[X]$, we have $P^{(t)}_{\mathsf{st}(s)}(E) = \overline{G}^{(t)}_s(\mathsf{st}^{-1}(E)\cap S)$.*

Proof. The proof uses Lemmas 4.2.17 and 4.2.18 and is similar to the proof of Theorem 3.3.23. □

In order to extend the result in Theorem 4.2.19 to all non-negative $t \in \mathbb{R}$, we follow the same path as Section 3.2. 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 3.2, we obtain the following result.

Theorem 4.2.20. *Suppose (VD), (OC) and (WF) hold. For any $s \in \mathrm{NS}(S)$, any $t \in \mathrm{NS}(T)$ and any $E \in \mathscr{B}[X]$, we have $P_{\mathsf{st}(s)}^{(\mathsf{st}(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$.*

Thus, in conclusion, we have the following theorem.

Theorem 4.2.21. *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 (VD), (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 $t \in \mathrm{NS}(T)$*

$$(\forall E \in \mathscr{B}[X])(P_{\mathsf{st}(s)}^{(\mathsf{st}(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)). \tag{4.2.23}$$

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 [1], 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.

4.2.2 A Weaker Markov Chain Ergodic Theorem

In Section 4.1, 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 4.1.4, we define an internal probability measure π' on $(S, \mathscr{I}(S))$ by letting $\pi'(\{s\}) = \frac{{}^*\pi(B(s))}{{}^*\pi(\bigcup_{s' \in S} B(s'))}$ for every $s \in S$. By Lemma 4.1.5, for any $A \in \mathscr{B}[X]$ 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 4.2.22. *Suppose (VD) and (WF) hold. There exists an infinite $t_0 \in T$ such that for every*

$A \in \mathscr{I}(S_1)$ and every $t \leq t_0$ we have

$$\pi'(A_S) \approx \sum_{i \in S} \pi'(i) G_i^{(t)}(A_S). \tag{4.2.24}$$

where A_S is the enlargement of A.

Proof. The proof is similar to the proof of Theorem 4.1.6. We use Theorem 4.2.14 instead of Theorem 3.3.16. □

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 4.2.23. *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 4.2.24. $\mathsf{st}^{-1}(O) \cap S \subset (^*O \cap S_1)_S \subset \mathsf{st}^{-1}(\overline{O}) \cap S$.

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 4.2.21.

Theorem 4.2.25. *Suppose (VD), (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 4.2.21, 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 4.2.24. □

We now show that $\{X_t'\}$ is strong regular for open balls.

Lemma 4.2.26. *Suppose (VD), (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 4.2.25, 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

$$T_O = \{t \in T : |G_{s_1}^{(t)}(({}^*O \cap S_1)_S) - G_{s_2}^{(t)}(({}^*O \cap S_1)_S)| < \frac{1}{t}\}. \tag{4.2.25}$$

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 $\{t_O : O \in \mathscr{B}\}$. □

By using essentially the same argument as in Theorem 3.1.19, we have the following result for $\{X_t'\}_{t \in T}$. The proof is omitted.

Theorem 4.2.27. *Suppose (VD), (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 4.2.22. Then for $\overline{\pi'}$-almost every $s \in S$ there exists an infinite $t' \in T$ such that*

$$G_s^{(t)}(({}^*O \cap S_1)_S) \approx \pi'(({}^*O \cap S_1)_S) \tag{4.2.26}$$

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 4.2.28. *Suppose (VD), (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 $\varepsilon > 0$ and a unbounded sequence of real numbers $\{k_n : n \in \mathbb{N}\}$ with $|g(x_0,k_n,O) - \pi(O)| > \varepsilon$ for all $n \in \mathbb{N}$. By Theorem 4.2.25 and Lemma 4.2.23, we have $|G_{s_0}^{(k_n)}(({}^*O \cap S_1)_S) - \pi'(({}^*O \cap S_1)_S)| > \varepsilon$ for all $n \in \mathbb{N}$. Let t' be the same infinite element in T as in Theorem 4.2.27. By overspill, there is an infinite $t_0 < t'$ such that $|G_{s_0}^{(t_0)}(({}^*O \cap S_1)_S) - \pi'(({}^*O \cap S_1)_S)| > \varepsilon$. 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 $|G_{s_0}^{(t_s)}(({}^*O \cap S_1)_S) - \pi'(({}^*O \cap S_1)_S)| > \varepsilon$. As $\overline{\pi'}(\mathsf{st}^{-1}(B) \cap S) = \pi(B)$, this contradicts with Theorem 4.2.27 hence completing the proof. ☐

We now generalize the convergence to all Borel sets. We will need the following definition.

Definition 4.2.29 ([41, Page. 85]). Let P_n and P be probability measures on a metric space X with Borel σ-algebra $\mathscr{B}[X]$. A subclass $\mathscr{C}$ of $\mathscr{B}[X]$ 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 $A \in \mathscr{C}$.

For separable metric spaces, we have the following result.

Lemma 4.2.30 ([34, Page. 416]). *Let P_n and P be probability measures on a separable metric space X with Borel σ-algebra $\mathscr{B}[X]$. A class $\mathscr{C}$ of Borel sets is a convergence determining class if $\mathscr{C}$ is closed under finite intersections and each open set in X is at most a countable union of elements in $\mathscr{C}$.*

Theorem 4.2.31. *Suppose (VD), (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 $P_x^{(t)}(.)$ weakly converges to $\pi(.)$.*

Proof. Let $\mathscr{B}'$ to be the smallest set containing $\mathscr{B}$ such that $\mathscr{B}'$ is closed under finite intersection. By Lemma 4.2.28, 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 4.2.30. ☐

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 4.1.16, we need $\{X'_t\}_{t\in T}$ to be strong regular on a larger class of sets.

Open Problem 4. *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 $A \in \mathscr{I}(S_1)$?*

Chapter 5

Push-down of Hyperfinite Markov Processes

In the previous chapter, we discuss how to construct hyperfinite Markov processes from standard Markov processes. The procedure for using hyperfinite Markov processes to construct standard Markov processes as well as stationary distributions is the reverse of the material that discussed in the previous chapter.

In Section 5.1, we begin with an internal probability measure on ${}^*X$ and use the standard part map to "push" the corresponding Loeb measure down to X to generate a standard probability measure. This push-down technique is useful in establishing existence result. We then 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 $x, y \in X$

$$\lim_{t\to\infty} \parallel P_x^{(t)}(\cdot) - P_y \parallel = 0. \tag{5.0.1}$$

Note that a Markov process with the merging property does not need to have a stationary distribution. In Section 5.2, we discuss conditions on $\{X_t\}_{t\geq 0}$ for it to have the merging property. Finally, we close with some remarks and open problems in Section 5.3.

5.1 Push-down Results

In Section 3.2, we discuss how to construct a corresponding hyperfinite Markov process for every s-tandard 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 5.1.1 ([11, Thm. 13.4.1])**.** *Let X be a Heine-Borel metric space equipped with Borel σ-algebra $\mathscr{B}[X]$. Let M be an internal probability measure defined on $({}^*X, {}^*\mathscr{B}[X])$. Let*

$$\mathscr{C} = \{C \subset X : \mathsf{st}^{-1}(C) \in \overline{{}^*\mathscr{B}[X]}\}. \tag{5.1.1}$$

Define a measure μ on the sets $\mathscr{C}$ 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 $\mathscr{C}$ is a σ-algebra. Obviously $\emptyset \in \mathscr{C}$. By Lemma 2.4.10, we know that $X \in \mathscr{C}$. We now show that it is closed under complement. Suppose $A \in \mathscr{C}$. It is easy to see that $\mathsf{st}^{-1}(A^c) = (\mathrm{NS}({}^*X) \setminus \mathsf{st}^{-1}(A))$. By Theorem 2.3.1 and the fact that $\overline{{}^*\mathscr{B}[X]}$ is a σ-algebra, $A^c \in \mathscr{C}$. We now show that $\mathscr{C}$ is closed under countable union. Suppose $\{A_i : i \in \mathbb{N}\}$ be a countable collection of pairwise disjoint elements from $\mathscr{C}$. 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 $\mathsf{st}^{-1}(A_i) \in \overline{{}^*\mathscr{B}[X]}$ for every $i \in \mathbb{N}$, we have $\mathsf{st}^{-1}(\bigcup_{i\in\omega} A_i) \in \overline{{}^*\mathscr{B}[X]}$. Hence $\bigcup_{i\in\omega} A_i \in \mathscr{C}$.

We now show that μ is a well-defined measure on $(X, \mathscr{C})$. Clearly $\mu(\emptyset) = 0$. Suppose $\{A_i\}_{i\in\omega}$ is a mutually disjoint collection from $\mathscr{C}$. We have

$$\mu(\bigcup_{i\in\omega} A_i) = \overline{M}(\mathsf{st}^{-1}(\bigcup_{i\in\omega} A_i)) = \overline{M}(\bigcup_{i\in\omega}(\mathsf{st}^{-1}(A_i))). \tag{5.1.2}$$

As A_i's are mutually disjoint, we know that $\mathsf{st}^{-1}(A_i)$'s are mutually disjoint. Thus,

$$\overline{M}(\bigcup_{i\in\omega}(\mathsf{st}^{-1}(A_i))) = \sum_{i\in\omega} \overline{M}(\mathsf{st}^{-1}(A_i)) = \sum_{i\in\omega} \mu(A_i). \tag{5.1.3}$$

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 2.3.1 and 2.3.9), we know that $\mathsf{st}^{-1}(B) \in \overline{{}^*\mathscr{B}[X]}$ for all $B \in \mathscr{B}[X]$. Consider any $B \in \mathscr{B}[X]$ 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 Lemma 8.1.1 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 Lemma 8.1.1.

Corollary 5.1.2. *Let X be a Heine-Borel metric space equipped with Borel σ-algebra $\mathscr{B}[X]$ and let S_X be a hyperfinite representation of X. Let M be an internal probability measure defined on $(S_X, \mathscr{I}[S_X])$. Let*

$$\mathscr{C} = \{C \subset X : \mathsf{st}^{-1}(C) \cap S_X \in \overline{\mathscr{I}[S_X]}\}. \tag{5.1.4}$$

Then the push-down measure M_p on the sets $\mathscr{C}$ 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 5.1.3. *Let X be a metric space equipped with Borel σ-algebra $\mathscr{B}[X]$, let ν be an internal probability measure on $(^*X, ^*\mathscr{B}[X])$ 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 $(X, \mathscr{B}[X])$. 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 $\mathscr{F}_n = \{F_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ forms a finite partition of X, and similarly for $\mathscr{G}_n = \{G_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ 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 2.4.10, $G_{n,k}$ 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 $\mathscr{F}_n$ (resp., $\mathscr{G}_n$). 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

Lemma 8.1.1, 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

$$\int_X \hat{f}_n \mathrm{d}\nu_p = \frac{k}{n}\nu_p(F_{n,k}) = \frac{k}{n}\overline{\nu}(G_{n,k}) = \int_{\mathrm{NS}(^*X)} \hat{g}_n \mathrm{d}\overline{\nu} \tag{5.1.5}$$

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. □

5.1.1 Construction of Standard Markov Processes

In Section 3.2, 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 $t_1, t_2 \in \mathrm{NS}(T)$:

$$(s_1 \approx s_2 \wedge t_1 \approx t_2) \implies (\forall A \in \mathscr{I}[S]\overline{G}_{s_1}^{(t_1)}(A) = \overline{G}_{s_2}^{(t_2)}(A)) \tag{5.1.6}$$

2.

$$(\forall s \in \mathrm{NS}(S))(\forall t \in \mathrm{NS}(T))(\overline{G}_s^{(t)}(\mathrm{NS}(S)) = 1). \tag{5.1.7}$$

For every $x \in X$, every $h \in \mathbb{R}^+$ and every $A \in \mathscr{B}[X]$, define

$$g(x,h,A) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(A) \cap S) \tag{5.1.8}$$

where $s \approx x$ and $t \approx h$. Such $g(x,h,A)$ is well-defined because of Eq. (5.1.6). By Lemma 8.1.1 and Eq. (5.1.7), it is easy to see that $g(x,h,.)$ is a probability measure on $(X, \mathscr{B}[X])$ 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 $(X, \mathscr{B}[X])$. We first recall Definition 2.2.15 and Theorem 2.2.16.

Definition 5.1.4. 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 5.1.5 ([3, 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 5.1.6. *Suppose $\{Y_t\}_{t\geq 0}$ satisfies Eqs. (5.1.6) and (5.1.7). Then for any $t_1,t_2\in \mathrm{NS}(T)$, any $s_0\in S$ and any $E\in\mathscr{B}[X]$, 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 $E\in\mathscr{B}[X]$. By Eqs. (5.1.6) and (5.1.7), 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 2.2.16 and the following claim.

Claim 5.1.7. *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 $G_{s_0}^{(t_1)}(\cdot)$ is an internal probability measure concentrating on a hyperfinite set, by Corollary 2.2.12, it is easy to see that ${}^*g(\cdot,\mathsf{st}(t_2),{}^*E)$ is S-integrable. As $g(\mathsf{st}(s),\mathsf{st}(t_2),E)=\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)$, 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 $\varepsilon\in\mathbb{R}^+$. Let $s_1\in S$ be any element such that $s_1\approx x_1$. Let $M=\{s\in S:(\forall A\in\mathscr{I}[S])(|G_s^{(t_2)}(A)-G_{s_1}^{(t_2)}(A)|<\varepsilon\}$. By Eq. (5.1.6), M contains every element in S which is infinitesimally close to s_1. By overspill, there is a $\delta\in\mathbb{R}^+$ such that

$$(\forall s\in S)({}^*d(s,s_1)<\delta\implies(\forall A\in\mathscr{I}[S])(|G_s^{(t_2)}(A)-G_{s_1}^{(t_2)}(A)|<\frac{\varepsilon}{2})). \tag{5.1.9}$$

This clearly implies that

$$(\forall s\in S)({}^*d(s,s_1)<\delta\implies(\forall E\in\mathscr{B}[X])(|\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)-\overline{G}_{s_1}^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)|<\varepsilon)). \tag{5.1.10}$$

By the construction of $g(\cdot,\mathsf{st}(t_2),E)$, we have $|g(x,\mathsf{st}(t_2),E)-g(x_1,\mathsf{st}(t_2),E)|<\varepsilon$ 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 $\overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S)$.

Lemma 5.1.8. *Suppose $\{Y_t\}_{t \in T}$ satisfies Eqs. (5.1.6) and (5.1.7). For any $t_1, t_2 \in \mathrm{NS}(T)$, $s_0 \in \mathrm{NS}(S)$ and $E \in \mathscr{B}[X]$, we have*

$$\overline{G}_{s_0}^{(t_1+t_2)}(\mathsf{st}^{-1}(E) \cap S) \approx \int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{5.1.11}$$

Proof. Pick some $E \in \mathscr{B}[X]$, some $s_0 \in \mathrm{NS}(S)$ and some $t_1, t_2 \in \mathrm{NS}(T)$. For any set $A \in \mathscr{I}[S]$ with $\mathsf{st}^{-1}(E) \cap S \subset A$, we have $\overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \leq \overline{G}_s^{(t_2)}(A)$. Hence we have

$$\int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \leq \int \overline{G}_s^{(t_2)}(A) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{5.1.12}$$

By Corollary 2.2.12, we have

$$\int \overline{G}_s^{(t_2)}(A) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) = \mathsf{st}\left(\int G_s^{(t_2)}(A) G_{s_0}^{(t_1)}(\mathrm{d}s)\right) = \mathsf{st}(G_{s_0}^{(t_1+t_2)}(A)) \tag{5.1.13}$$

Hence, we have

$$\int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \leq \inf\{\mathsf{st}(G_{s_0}^{(t_1+t_2)}(A)) : \mathsf{st}^{-1}(E) \cap S \subset A \in \mathscr{I}[S]\}. \tag{5.1.14}$$

Similarly, we have

$$\int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \geq \sup\{\mathsf{st}(G_{s_0}^{(t_1+t_2)}(B)) : \mathsf{st}^{-1}(E) \cap S \supset B \in \mathscr{I}[S]\}. \tag{5.1.15}$$

Hence, by the construction of Loeb measure, we have

$$\overline{G}_{s_0}^{(t_1+t_2)}(\mathsf{st}^{-1}(E) \cap S) \approx \int \overline{G}_s^{(t_2)}(\mathsf{st}^{-1}(E) \cap S) \overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{5.1.16}$$

□

We now establish the main result of this section.

Theorem 5.1.9. *Suppose $\{Y_t\}_{t \in T}$ satisfies Eqs. (5.1.6) and (5.1.7). Then for any $h_1, h_2 \in \mathbb{R}^+$, any $x_0 \in X$ and any $E \in \mathscr{B}[X]$ we have*

$$g(x_0, h_1 + h_2, E) = \int g(x, h_2, E) g(x_0, h_1, \mathrm{d}x). \tag{5.1.17}$$

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 $E\in\mathscr{B}[X]$. 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 5.1.8, we have

$$g(x_0,h_1+h_2,E)=\overline{G}_{s_0}^{(t_1+t_2)}(\mathsf{st}^{-1}(E)\cap S)=\int\overline{G}_{s}^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s). \tag{5.1.18}$$

By Eq. (5.1.6), 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 5.1.7, 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 8.1.5, we have

$$\int_S\overline{G}_{s}^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{5.1.19}$$
$$=\int_{\mathrm{NS}(S)}\overline{G}_{s}^{(t_2)}(\mathsf{st}^{-1}(E)\cap S)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{5.1.20}$$
$$=\int_{\mathrm{NS}(S)}\mathsf{st}({}^*g(s,h_2,{}^*E))\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{5.1.21}$$
$$=\int_{\mathrm{NS}(S)}g(\mathsf{st}(s),h_2,E)\overline{G}_{s_0}^{(t_1)}(\mathrm{d}s) \tag{5.1.22}$$
$$=\int_X g(x,h_2,E)g(x_0,h_1,\mathrm{d}x). \tag{5.1.23}$$

Note that the last step follows from Lemma 8.1.5. 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 $\mathscr{B}[X]$. 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 $\{Y_t\}_{t\in T}$.

5.1.2 Push down of Weakly Stationary Distributions

Recall from Definition 3.1.5 that an internal probability measure π on $(S,\mathscr{I}[S])$ is a weakly stationary distribution if there is an infinite t_0 such that

$$(\forall t\leq t_0)(\forall A\in\mathscr{I}(S))(\pi(A)\approx\sum_{i\in S}\pi(\{i\})p^{(t)}(i,A)) \tag{5.1.24}$$

$p^{(t)}(i,A)$ denote the t-step internal transition probability of a hyperfinite Markov process.

In Section 5.1.1, 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. (5.1.6) and (5.1.7) 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 5.1.1. 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 Lemma 8.1.1. It is easy to see that Π_p is a probability measure on $(X,\mathscr{B}[X])$.

We first establish the following lemma.

Lemma 5.1.10. *Suppose $\{Y_t\}_{t\geq 0}$ satisfies Eqs.* (5.1.6) *and* (5.1.7). *Then for any $t\in\mathrm{NS}(T)$ and any $E\in\mathscr{B}[X]$, 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 5.1.6. □

Lemma 5.1.11. *Suppose $\{Y_t\}_{t\in T}$ satisfies Eqs.* (5.1.6) *and* (5.1.7). *Then for any $t\in\mathrm{NS}(T)$ and any $E\in\mathscr{B}[X]$, we have*

$$\overline{\Pi}(\mathsf{st}^{-1}(E)\cap S)\approx\int G_s^{(t)}(\mathsf{st}^{-1}(E)\cap S)\overline{\Pi}(\mathrm{d}s). \tag{5.1.25}$$

Proof. The proof is similar to Lemma 5.1.8 □

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 5.1.12. *Suppose $\{Y_t\}_{t\geq 0}$ satisfies Eqs.* (5.1.6) *and* (5.1.7). *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 Lemma 8.1.1 and the fact that $\overline{\Pi}(\mathrm{NS}(S))=1$, we know that Π_p is a probability measure on $(X,\mathscr{B}[X])$.

Fix $t_0 \in \mathbb{R}^+$ and $A \in \mathscr{B}[X]$. 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 5.1.11, we have

$$\Pi_p(A) = \overline{\Pi}(\mathsf{st}^{-1}(A) \cap S) = \int G_s^{(t)}(\mathsf{st}^{-1}(A) \cap S)\overline{\Pi}(\mathrm{d}s). \tag{5.1.26}$$

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 5.1.9, we have

$$\int_S \overline{G}_s^{(t)}(\mathsf{st}^{-1}(A) \cap S)\overline{\Pi}(\mathrm{d}s) \tag{5.1.27}$$

$$= \int_{\mathrm{NS}(S)} \mathsf{st}({}^*g(s,t_0,{}^*A))\overline{\Pi}(\mathrm{d}s) \tag{5.1.28}$$

$$= \int_X g(x,t_0,A)\Pi_p(\mathrm{d}x). \tag{5.1.29}$$

Hence completing the proof. □

Suppose we start with a standard Markov process $\{X_t\}_{t\geq 0}$ satisfying (VD), (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. (5.1.6) and (5.1.7). 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.

5.1.3 Existence of Stationary Distributions

The existence of stationary distribution for discrete-time Markov processes with finite state space is well-understood (e.g [42, 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 5.1.13. A $n \times n$ matrix $\mathbb{P}$ is regular if some power of $\mathbb{P}$ has only positive entries.

Theorem 5.1.14. *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 5.1.15. 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 5.1.16. *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 5.1.17. 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 5.1.18. *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 5.1.14. □

Note that if Π is *stationary then Π is weakly stationary as in Definition 3.1.5. The following theorem gives a sufficient condition for regularity of $\mathbb{P}$.

Theorem 5.1.19. *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 5.1.16.

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 5.1.20. *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. (4.2.23). 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 5.1.18, 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 5.1.12, 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 5.1.20 provides a way to look for stationary distributions.

Example 5.1.21 (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 3.3.31. 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 $G^{(\delta t)}_{s_1}(\{s_2\}) > 0$ for all $s_1, s_2 \in S$). By Theorem 5.1.18, 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 *stationary 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.

5.2 Merging of Markov Processes

In Section 4.1, 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 (VD), (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 5.2.1. A Markov process $\{X_t\}_{t\geq 0}$ has the merging property if for every two points $x, y \in X$, we have

$$\lim_{t\to\infty} \parallel P^{(t)}_x(.) - P^{(t)}_y(.) \parallel = 0 \tag{5.2.1}$$

where $P_x^{(t)}(.)$ denotes the transition measure and $\parallel P_x^{(t)}(.) - P_y^{(t)}(.) \parallel$ denotes the total variation distance between $P_x^{(t)}(.)$ and $P_y^{(t)}(.)$.

Saloff-Coste and Zúñiga [44] 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 [12].

In this section, we give sufficient conditions to ensure that Markov processes have the merging property. The following definition is analogous to Definition 3.1.11.

Definition 5.2.2. 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 $n \in \mathbb{N}$

$$Q_{(x_1,x_2)}(\exists t\ Z_t \in U(y_1, \frac{1}{n}) \times U(y_2, \frac{1}{n})) = 1. \tag{5.2.2}$$

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 $\varepsilon_0 \cdot (\frac{t}{\delta t}) \approx 0$ for all $t \in T$. As in Section 3.3, 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 2.1.12 and Theorem 2.4.6, 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 3.2.12 and are denoted by $\{G_i^{(t)}(\cdot)\}_{i\in S, t\in T}$.

Lemma 5.2.3. *Suppose $\{X_t\}_{t\geq 0}$ satisfies (VD), (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 (VD), (SF) and (OC), by Theorem 3.3.31, we have

$$P_{\mathsf{st}(s)}^{(\mathsf{st}(t))}(E) = \overline{G}_s^{(t)}(\mathsf{st}^{-1}(E) \cap S) \tag{5.2.3}$$

hence implies that $\{X'_t\}_{t\in T}$ satisfies Eqs. (5.1.6) and (5.1.7). Let $\{X_t^p\}_{t\geq 0}$ denote the standard Markov process obtained from pushing down $\{X'_t\}_{t\in T}$ as in Section 5.1.1. 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 $E \in \mathscr{B}[X]$ 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 5.2.2, we know that $Q_{(x_1,x_2)}(\exists t > 0\, Z_t \in U_1 \times U_2) = 1$. This implies that

$$q_{(x_1,x_2)}(\exists t > 0\, Z_t^p \in U_1 \times U_2) = 1 \tag{5.2.4}$$

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

$$\overline{F}_{(x_1,x_2)}(\exists t \in \mathrm{NS}(T)\, Z_t' \in (\mathsf{st}^{-1}(U_1) \times \mathsf{st}^{-1}(U_2)) \cap (S \times S)) = 1 \tag{5.2.5}$$

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. □

The proof of the following theorem is similar to the proof of Theorem 3.1.19.

Theorem 5.2.4. *Suppose $\{X_t\}_{t \geq 0}$ satisfies (VD), (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 $A \in {}^*\mathscr{B}[X]$ 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

$$M = \{\omega : \exists t < t_0 - 1, X_t'(\omega) \approx Y_t(\omega) \approx y\}. \tag{5.2.6}$$

By Lemma 5.2.3, for all $n \in \mathbb{N}$, we have

$$\overline{F}_{(x_1,x_2)}(\exists t \in \mathrm{NS}(T)\, Z_t' \in ({}^*U(y, \frac{1}{n}) \times {}^*U(y, \frac{1}{n})) \cap (S \times S)) = 1. \tag{5.2.7}$$

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 3.1.8, we know that $\overline{F}_{(x_1,x_2)}(M)=1$. By Theorem 3.3.21, we know that $\{X'_t\}_{t\in T}$ is strong regular. Thus we have:

$$|\overline{G}_{x_1}^{(t_0)}(A)-\overline{G}_j^{(t)}(A)| \tag{5.2.8}$$

$$=|\overline{F}_{(x_1,x_2)}(X'_{t_0}\in A)-\overline{F}_{(x_1,x_2)}(Y_{t_0}\in A)| \tag{5.2.9}$$

$$=|\overline{F}_{(x_1,x_2)}((X'_{t_0}\in A)\cap M)-\overline{F}_{(x_1,x_2)}((Y_{t_0}\in A)\cap M)| \tag{5.2.10}$$

$$=0. \tag{5.2.11}$$

□

We now establish the following merging result for the standard Markov process $\{X_t\}_{t\geq 0}$.

Theorem 5.2.5. *Suppose $\{X_t\}_{t\geq 0}$ satisfies (VD), (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 $\varepsilon>0$ and fix two standard $x_1,x_2\in X$. By Theorem 5.2.4, we know that $|G_{x_1}^{(t)}(A)-G_{x_2}^{(t)}(A)|<\varepsilon$ for all infinite $t\in T$ and all $A\in {}^*\mathscr{B}[X]$. Let $M=\{t\in T:(\forall A\in {}^*\mathscr{B}[X])(|G_{x_1}^{(t)}(A)-G_{x_2}^{(t)}(A)|<\varepsilon)\}$. By the underspill principle, there exists a $t_0\in \mathrm{NS}(T)$ such that $|G_{x_1}^{(t_0)}(A)-G_{x_2}^{(t_0)}(A)|<\varepsilon$ for all $A\in {}^*\mathscr{B}[X]$. Pick a standard $t_1>t_0$ and let $t_2\in T$ be the first element greater than t_1.

Claim 5.2.6. *$|G_{x_1}^{(t_2)}(A)-G_{x_2}^{(t_2)}(A)|<\varepsilon$ for all $A\in {}^*\mathscr{B}[X]$.*

Proof. Pick $t_3\in T$ such that $t_0+t_3=t_2$ and any $A\in {}^*\mathscr{B}[X]$. Then we have

$$|G_{x_1}^{(t_2)}(A)-G_{x_2}^{(t_2)}(A)| \tag{5.2.12}$$

$$\approx |\sum_{y\in S}G_{x_1}^{(t_1)}(\{y\})G_y^{(t_2)}(A)-\sum_{y\in S}G_{x_2}^{(t_1)}(\{y\})G_y^{(t_2)}(A)| \tag{5.2.13}$$

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 3.1.24, we know that

$$|G_{x_1}^{(t_2)}(A)-G_{x_2}^{(t_2)}(A)|\lessapprox \| G_{x_1}^{(t_1)}(.)-G_{x_2}^{(t_1)}(.)\|. \tag{5.2.14}$$

Since this is true for all internal A, we have established the claim. □

By the construction of Loeb measure, we know that

$$(\forall B \in \mathscr{B}[X])(|\overline{G}_{x_1}^{(t_2)}(\mathsf{st}^{-1}(B)\cap S) - \overline{G}_{x_2}^{(t_2)}(\mathsf{st}^{-1}(B)\cap S)| < \varepsilon). \tag{5.2.15}$$

By Theorem 3.3.31 and the fact that $t_2 \approx t_1$, we know that $|P_{x_1}^{(t_1)}(B) - P_{x_2}^{(t_1)}(B)| < \varepsilon$ for all $B \in \mathscr{B}[X]$. This shows that $\{X_t\}_{t\geq 0}$ has the merging property. □

5.3 Remarks and Open Problems

(i) So far we have required that the state space X is a metric space satisfying the Heine-Borel property. Such an X is automatically a σ-compact locally compact metric space. Let $X = \bigcup_{n\in\mathbb{N}} K_n$ where every K_n is a compact subset of X. Heine-Borel property is essential since it implies that ${}^*X = \bigcup_{n\in\mathbb{N}} {}^*K_n$. However, the Heine-Borel condition turns out to be quite strong. For example, $(0,1)$ and set of rational numbers $\mathbb{Q}$, while they are σ-compact and locally compact spaces, do not satisfy the Heine-Borel property. The following theorem shows that, for every σ-compact locally compact metric space, we can impose a Heine-Borel metric d_H on X without changing the topology on X.

Theorem 5.3.1. *Let (X,d) be σ-compact 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. 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

$$d_H(x,y) = d(x,y) + |f(x) - f(y)|. \tag{5.3.1}$$

Then

$$d_H(x,z) = d(x,z) + |f(x) - f(z)| \leq d(x,y) + |f(x) - f(y)| + d(y,z) + |f(y) - f(z)| \tag{5.3.2}$$

hence d_H is a metric on X.

Claim 5.3.2. *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 5.3.3. *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. □

□

As the Heine-Borel metric induces the same topology in X, instead of assuming the state space X satisfies the Heine-Borel condition we only need X to be a σ-compact locally compact space.

(ii) 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 $A \in {}^*\mathscr{B}[X]$ 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 [27], Loeb showed that, for any Hausdorff topological space X, there is a hyperfinite partition $\mathscr{B}_F$ 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 $\{A_i : i \leq N\} \in \mathscr{P}({}^*\mathscr{B}[X])$ 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

$$\sup_{x\in A_i} {}^*f(x) - \inf_{x\in A_i} {}^*f(x) \approx 0 \tag{5.3.3}$$

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 $A \in \mathscr{B}[X]$ 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 $B \in {}^*\mathscr{B}[X]$. 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.

(iii) It is possible to weaken the conditions mentioned in the Markov chain ergodic theorem (Theorem 4.1.16). In particular, it would be interesting to reduce (SF) to (WF). In Section 4.2, 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 $A \in \mathscr{I}[S]$, any $i, j \in \mathrm{NS}(S)$ and any $t \in T$ we have:

$$(i \approx j) \implies (G_x^{(t)}(A) \approx G_y^{(t)}(A)). \tag{5.3.4}$$

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 $B \in {}^*\mathscr{B}[X]$, 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 $A \in \mathscr{B}[X]$. Thus, an affirmative answer to Open Problem 5 should allow us to reduce (SF) to (WF) in the Markov chain ergodic theorem (Theorem 4.1.16).

(iv) The following nonstandard measure theoretical question is related to the previous point. Let X be a topological space and let $(X, \mathscr{B}[X])$ be a Borel-measurable space. The question is: is an internal probability measure on $({}^*X, {}^*\mathscr{B}[X])$ determined by its value on $\{{}^*A : A \in \mathscr{B}[X]\}$? For nonstandard extensions of standard probability measures on $(X, \mathscr{B}[X])$, the answer is affirmative by the transfer principle. For general internal probability measures on $({}^*X, {}^*\mathscr{B}[X])$, 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 5. *Let X be a topological space and let $(X,\mathscr{B}[X])$ be a Borel-measurable space. Let P be a probability measure on $(X,\mathscr{B}[X])$ and let P_1 be an internal probability measure on $({}^*X,{}^*\mathscr{B}[X])$. Suppose $P_1({}^*A)\approx {}^*P({}^*A)$ for all $A\in\mathscr{B}[X]$, is it true that $\overline{P_1}=\overline{{}^*P}$?*

We do have the following partial result.

Lemma 5.3.4. *Let us consider $([0,1],\mathscr{B}[[0,1]])$ and let P be a probability measure on it. Let P_1 be an internal probability measure on $({}^*[0,1],{}^*\mathscr{B}[[0,1]])$ such that $P_1({}^*A)\approx {}^*P({}^*A)$ for all $A\in\mathscr{B}[[0,1]]$. 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 5 may follow from a variation of Theorem 3.3.33.

(v) 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 5.1.20, 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 3.1.26, under those conditions in Theorem 4.1.16, 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 6. *Is there an example of a hyperfinite Markov process where its *stationary distribution differs from some of its weakly stationary distribution?*

Open Problem 7. *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 3.1.26?*

(vi) In Section 4.2.2, 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 5.3.5. *Let* $(X, \mathcal{T})$ *be a topological space and let* $(X, \mathscr{B}[X])$ *be a Borel-measurable space. Let* $\{P_n : n \in \mathbb{N}\}$ *and* P *be Radon probability measures on* $(X, \mathscr{B}[X])$*. Suppose*

$$\lim_{n\to\infty} \sup_{U\in\mathcal{T}} |P_n(U) - P(U)| = 0. \tag{5.3.5}$$

Then $(P_n : n \in \mathbb{N})$ *converges to* P *in total variation distance.*

Proof. Pick $\varepsilon > 0$. There is a $n_0 \in \mathbb{N}$ such that $\sup_{U\in\mathcal{T}} |P_n(U) - P(U)| < \frac{\varepsilon}{4}$ for all $n > n_0$. Let $\mathcal{K}(X)$ denote the collection of compact subsets of X. Then we have $\sup_{K\in\mathcal{K}(X)} |P_n(K) - P(K)| < \frac{\varepsilon}{4}$ for all $n > n_0$. Fix $B \in \mathscr{B}[X]$ 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 $P_{n_1}(U) - P_{n_1}(K) < \frac{\varepsilon}{4}$. We then have

$$\begin{aligned}
&|P_{n_1}(B) - P(B)| && (5.3.6)\\
&\leq |P_{n_1}(U) - P(K)| && (5.3.7)\\
&\leq |P_{n_1}(U) - P_{n_1}(K)| + |P_{n_1}(K) - P(K)| && (5.3.8)\\
&\leq \frac{\varepsilon}{2}. && (5.3.9)
\end{aligned}$$

This implies that $\sup_{B\in\mathscr{B}[X]} |P_{n_1}(B) - P(B)| < \varepsilon$. 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.

(vii) 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 5.3.6. 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 $A \subset X$, $t_0 > 0$, $0 < \varepsilon < 1$, 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.
- $P_x^{(t_0)}(B) > \varepsilon\mu(B)$ 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 [32]. Note that the skeleton condition is usually hard to check.

Theorem 5.3.7 ([32, 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 5.3.8 ([39, Example. 3])**.** Let $X = \{1, 2, \dots\}$. 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

$$P[(\forall n)(X_n = x+n) | X_0 = x] = \prod_{i=x}^{\infty}(1 - \frac{1}{i^2}) = \frac{x-1}{x} > 0 \tag{5.3.10}$$

Hence the convergence only holds if we start at $\{1\}$.

The Markov chain ergodic theorem developed in this paper (Theorem 4.1.16) 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.

Chapter 6

Introduction to Statistical Decision Theory

More than eighty years after its formulation, statistical decision theory has served as a rigorous foundation of statistics. One of the most fundamental problems in statistical decision theory, known as the complete class theorem, is to study the relation between frequentist and Bayesian optimality. There is a long line of research, originating with Wald's development of statistical decision theory [54–57], that connects frequentist and Bayesian optimality [5, 9, 10, 20, 25, 43, 50, 52–57]. One of the key results, due to Le Cam [25], building off work of Wald, can be summarized as follows: under some technical conditions, every admissible procedure is a limit of Bayes procedures.

This and related results deepen our understanding of both frequentist and Bayesian optimality. In one direction, optimal frequentist procedures have (quasi) Bayesian interpretations that often provide insight into strengths and weaknesses from an average-case perspective. In the other direction, optimal frequentist procedures can be constructed via Bayes' rule from carefully chosen priors or generalized priors, such as improper priors or sequences thereof.

We give a general overview of statistical decision theory as well as an extensive literature review on complete class theorems in this chapter. In Section 6.1, we introduce basic notions and key results in standard statistical decision theory: domination, admissibility, and its variants; Bayes optimality; and basic complete class and essentially complete class results. Classic treatments can be found in [14] and [8], the latter emphasizing the connection with game theory, but restricting itself to finite discrete spaces. A modern treatment can be found in [26].

In Section 6.2, we give a summary of extensive literature on complete class theorems. For finite parameter spaces, it is well-known that a decision procedure is extended admissible if and only if it is Bayes. We shall see that various relaxations of this classical equivalence have been established for infinite parameter spaces, but these extensions are each subject to technical conditions that limit their applicability, especially to modern (semi- and nonparametric) statistical problems.

6.1 Standard Preliminaries

A (non-sequential) statistical decision problem is defined in terms of a *parameter* space Θ, each element of which represents a possible state of nature; a set $\mathbb{A}$ of *actions* available to the statistician; a function $\ell : \Theta \times \mathbb{A} \to \mathbb{R}_{\geq 0}$ characterizing the *loss* associated with taking action $a \in \mathbb{A}$ in state $\theta \in \Theta$; and finally, a family $P = (P_\theta)_{\theta \in \Theta}$ of probability measures on a measurable *sample* space X. On the basis of an observation from P_θ for some unknown element $\theta \in \Theta$, the statistician decides to take a (potentially randomized) action a, and then suffers the loss $\ell(\theta, a)$.

Formally, having fixed a σ-algebra on the space $\mathbb{A}$ of actions, every possible response by the statistician is captured by a *(randomized) decision procedure*, i.e., a map δ from X to the space $\mathscr{M}_1(\mathbb{A})$ of probability measures $\mathbb{A}$. As is customary, we will write $\delta(x,A)$ for $(\delta(x))(A)$. The expected loss, or *risk*, to the statistician in state θ associated with following a decision procedure δ is

$$r_\delta(\theta) = r(\theta, \delta) = \int_X \Big[\int_{\mathbb{A}} \ell(\theta, a) \delta(x, \mathrm{d}a) \Big] P_\theta(\mathrm{d}x). \tag{6.1.1}$$

For the risk function to be well-defined, the maps $x \mapsto \int_{\mathbb{A}} \ell(\theta, a)\delta(x, \mathrm{d}a)$, for $\theta \in \Theta$, must be measurable, and so we will restrict our attention to those decision procedures satisfying this weak measurability criterion. A decision procedure δ is said to have finite risk if $r_\delta(\theta) \in \mathbb{R}$ for all $\theta \in \Theta$. Let $\mathscr{D}$ denote the set of randomized decision procedures with finite risk.

The set $\mathscr{D}$ may be viewed as a convex subset of a vector space. In particular, for all $\delta_1, \dots, \delta_n \in \mathscr{D}$ and $p_1, \dots, p_n \in \mathbb{R}_{\geq 0}$ with $\sum_i p_i = 1$, define $\sum_i p_i \delta_i : X \to \mathscr{M}_1(\mathbb{A})$ by $(\sum_i p_i \delta_i)(x) = \sum_i p_i \delta_i(x)$ for $x \in X$. Then $r(\theta, \sum_i p_i \delta_i) = \sum_i p_i r(\theta, \delta_i) < \infty$, and so we see that $\sum_i p_i \delta_i \in \mathscr{D}$ and $r(\theta, \cdot)$ is a linear function on $\mathscr{D}$ for every $\theta \in \Theta$. For a subset $D \subseteq \mathscr{D}$, let $\mathrm{conv}(D)$ denote the set of all finite convex combinations of decision procedures $\delta \in D$.

A decision procedure $\delta \in \mathscr{D}$ is called *nonrandomized* if, for all $x \in X$, there exists $d(x) \in \mathbb{A}$ such that $\delta(x, A) = 1$ if and only if $d(x) \in A$, for all measurable sets $A \subseteq \mathbb{A}$. Let $\mathscr{D}_0 \subseteq \mathscr{D}$ denote the subset of all nonrandomized decision procedures. Under mild measurability assumptions, every $\delta \in \mathscr{D}_0$ can be associated with a map $x \mapsto d(x)$ from X to $\mathbb{A}$ for which the risk satisfies

$$r(\theta,\delta) = \int_X \ell(\theta, d(x)) P_\theta(\mathrm{d}x). \tag{6.1.2}$$

Finally, writing $S^{[<\infty]}$ for the set of all finite subsets of a set S, let

$$\mathscr{D}_{0,FC} = \bigcup_{D \in \mathscr{D}_0^{[<\infty]}} \mathrm{conv}(D) \tag{6.1.3}$$

be the set of randomized decision procedures that are finite convex combinations of nonrandomized decision procedures. Note that $\mathscr{D}_0 \subset \mathscr{D}_{0,FC} \subset \mathscr{D}$ and $\mathscr{D}_{0,FC}$ is convex.

6.1.1 Admissibility

In general, the risk functions of two decision procedures are incomparable, as one procedure may present greater risk in one state, yet less risk in another. Some cases, however, are clear cut: the notion of domination induces a partial order on the space of decision procedures.

Definition 6.1.1. Let $\varepsilon \geq 0$ and $\delta, \delta' \in \mathscr{D}$. Then δ is ε-dominated by δ' if

1. $\forall \theta \in \Theta \; r(\theta,\delta') \leq r(\theta,\delta) - \varepsilon$, and

2. $\exists \theta \in \Theta \; r(\theta,\delta') \neq r(\theta,\delta)$.

Note that δ is *dominated* by δ' if δ is 0-dominated by δ'. If a decision procedure δ is ε-dominated by another decision procedure δ', then, computational issues notwithstanding, δ should be eliminated from consideration. This gives rise to the following definition:

Definition 6.1.2. Let $\varepsilon \geq 0$, $\mathscr{C} \subseteq \mathscr{D}$, and $\delta \in \mathscr{D}$.

1. δ is *ε-admissible among $\mathscr{C}$* unless δ is ε-dominated by some $\delta' \in \mathscr{C}$.

2. δ is *extended admissible among $\mathscr{C}$* if δ is ε-admissible among $\mathscr{C}$ for all $\varepsilon > 0$.

Again, note that δ is *admissible among $\mathscr{C}$* if δ is 0-admissible among $\mathscr{C}$. Clearly admissibility implies extended admissibility. In other words, the class of all extended admissible decision procedures contains the class of all admissible decision procedures.

Admissibility leads to the notion of a complete class.

Definition 6.1.3. Let $\mathscr{A}, \mathscr{C} \subseteq \mathscr{D}$. Then $\mathscr{A}$ is a *complete* subclass of $\mathscr{C}$ if, for all $\delta \in \mathscr{C} \setminus \mathscr{A}$, there exists $\delta_0 \in \mathscr{A}$ such that δ_0 dominates δ. Similarly, $\mathscr{A}$ is an *essentially complete* subclass of $\mathscr{C}$ if, for all $\delta \in \mathscr{C} \setminus \mathscr{A}$, there exists $\delta_0 \in \mathscr{A}$ such that $r(\theta, \delta_0) \leq r(\theta, \delta)$ for all $\theta \in \Theta$. An *essentially complete class* is an essentially complete subclass of $\mathscr{D}$.

If a decision procedure δ is admissible among $\mathscr{C}$, then every complete subclass of $\mathscr{C}$ must contain δ. Note that the term *complete class* is usually used to refer to a complete subclass of some essentially complete class (such as $\mathscr{D}$ itself or $\mathscr{D}_0$ under the conditions described in Section 6.1.3.)

The next lemma captures a key consequence of essential completeness:

Lemma 6.1.4. *Suppose $\mathscr{A}$ is an essentially complete subclass of $\mathscr{C}$, then extended admissible among $\mathscr{A}$ implies extended admissible among $\mathscr{C}$.*

The class of extended admissible estimators plays a central role in this paper. It is not hard, however, to construct statistical decision problems for which the class is empty, and thus not a complete class.

Example 6.1.5. Consider a statistical decision problem with sample space $X = \{0\}$, parameter space $\Theta = \{0\}$, action space $\mathbb{A} = (0,1]$, and loss function $\ell(0,d) = d$. Then every decision procedure is a constant function, taking some value in $\mathbb{A}$. For all $c \in (0,1]$, the procedure $\delta \equiv c$ is $c/2$-dominated by the decision procedure $\delta' \equiv c/2$. Hence, there is no extended admissible estimator, hence the extended admissible procedures do not form a complete class.

The following result gives conditions under which the class of extended admissible estimators are a complete class. (See [8, §5.4–5.6 and Thm. 5.6.3] and [14, §2.6 Cor. 1] for related results for finite spaces.)

Theorem 6.1.6. *Let $\mathscr{C} \subseteq \mathscr{D}$. Suppose that, for all sequences $\delta, \delta_1, \delta_2, \ldots \in \mathscr{C}$ and non-decreasing sequences $\varepsilon_1, \varepsilon_2, \cdots \in \mathbb{R}_{>0}$ such that $\varepsilon_0 = \lim_i \varepsilon_i$ exists and δ is ε_i-dominated by δ_i for all $i \in \mathbb{N}$, there is a decision procedure $\delta_0 \in \mathscr{C}$ such that δ is ε_0-dominated by δ_0. Then the set of procedures that are extended admissible among $\mathscr{C}$ form a complete subclass of $\mathscr{C}$.*

Proof. Let $\mathscr{S} = \{x \in \mathbb{R}^\Theta : (\exists \delta \in \mathscr{C})\, (\forall \theta \in \Theta)\, x(\theta) = r(\theta, \delta)\}$ denote the risk set of $\mathscr{C}$. Pick $\delta \in \mathscr{C}$ and suppose δ is not extended admissible among $\mathscr{C}$. Let

$$Q_\varepsilon(\delta) = \{x \in \mathbb{R}^\Theta : (\forall \theta \in \Theta)(x(\theta) \leq r(\theta, \delta) - \varepsilon)\}. \tag{6.1.4}$$

Let M be the set $\{\varepsilon \in \mathbb{R}_{>0} : Q_\varepsilon(\delta) \cap \mathscr{S} \neq \emptyset\}$, which is nonempty because δ is not extended admissible among $\mathscr{C}$. As the risk is nonnegative and finite, M is also bounded above. Hence there exists a least upper bound ε_0 of M. Pick a non-decreasing sequence $\varepsilon_1, \varepsilon_2, \ldots \in M$ that converges to ε_0. We now construct a (potentially infinite) sequence of decision procedures inductively:

1. Choose $\delta_1 \in \mathscr{C}$ such that δ is ε_1-dominated by δ_1. Because M is nonempty, there must exist such a procedure.

2. Suppose we have chosen $\delta_1, \ldots, \delta_i \in \mathscr{C}$, and suppose there is an index $j \in \mathbb{N}$ such that δ is ε_j-dominated by δ_i but δ is not ε_{j+1}-dominated by δ_i. Then we choose $\delta_{i+1} \in \mathscr{C}$ such that δ is ε_{j+1}-dominated by δ_{i+1}. Because M contains ε_{j+1}, there must exist such a procedure. If no such index j exist, the process halts at stage i.

Suppose the process halts at some finite stage i_0. Then for, all $j \in \mathbb{N}$, δ is not ε_j-dominated by δ_{i_0} or δ is ε_{j+1}-dominated by δ_{i_0}. But δ is ε_1-dominated by δ_{i_0} and so, by induction, δ is ε_j-dominated by δ_{i_0} for all $j \in \mathbb{N}$. As the sequence $\varepsilon_1, \varepsilon_2, \ldots$ is non-decreasing and has a limit ε_0, it follows easily via a contrapositive argument that δ is even ε_0-dominated by δ_{i_0}. If δ_{i_0} were not extended admissible among $\mathscr{C}$, then this would contradict the fact that ε_0 is a least upper bound on M.

Now suppose the process continues indefinitely. Then the claim is that δ is ε_i-dominated by δ_i for all $i \in \mathbb{N}$. Clearly this holds for $i = 1$. Supposing it holds for $i \le k$. Then δ is ε_i-dominated by δ_k for all $i \le k$ and there exists $j \in \mathbb{N}$ such that δ is ε_j-dominated by δ_k but δ is not ε_{j+1}-dominated by δ_k. It follows that $j \ge k$, hence δ is ε_{k+1}-dominated by δ_{k+1}, as was to be shown.

Thus, by hypothesis, there is a decision procedure $\delta' \in \mathscr{C}$ such that δ is ε_0-dominated by δ'. As ε_0 is the least upper bound of M, δ' is also extended admissible among $\mathscr{C}$, completing the proof. □

6.1.2 Bayes Optimality

Consider now the Bayesian framework, in which one adopts a *prior*, i.e., a probability measure π defined on some σ-algebra on Θ. Irrespective of the interpretation of π, we may define the *Bayes* risk of a procedure as the expected risk under a parameter chosen at random from π.[1]

Definition 6.1.7. Let $\delta \in \mathscr{D}$, $\varepsilon \ge 0$, and $\mathscr{C} \subseteq \mathscr{D}$, and let π_0 be a prior.

1. The *Bayes risk under π_0 of δ* is $r(\pi_0, \delta) = \int_\Theta r(\theta, \delta)\pi_0(\mathrm{d}\theta)$.

[1] We must now also assume that $r(\cdot, \delta)$ is a measurable function for every $\delta \in \mathscr{D}$. Normally, there is a natural choice of σ-algebra on Θ that satisfies this constraint. Even if there is no natural choice, there is always a sufficiently rich σ-algebra that renders every risk function measurable. In particular, the power set of Θ suffices. Note that the σ-algebra determines the set of possible prior distributions. In the extreme case where the σ-algebra on Θ is taken to be the entire power set, the set of prior distributions contain the purely atomic distributions and these are the only distributions if and only if there is no real-valued measurable cardinal less than or equal to the continuum [19, Thm. 1D]. As we will see, the purely atomic distributions suffice to give our complete class theorems.

2. δ is *ε-Bayes under π_0 among $\mathscr{C}$* if $r(\pi_0,\delta)<\infty$ and, for all $\delta'\in\mathscr{C}$, we have $r(\pi_0,\delta)\le r(\pi_0,\delta')+\varepsilon$.

3. δ is *Bayes under π_0 among $\mathscr{C}$* if δ is 0-Bayes under π_0 among $\mathscr{C}$.

4. δ is *extended Bayes among $\mathscr{C}$* if, for all $\varepsilon>0$, there exists a prior π such that δ is ε-Bayes under π among $\mathscr{C}$.

5. δ is *ε-Bayes among $\mathscr{C}$* (resp., *Bayes among $\mathscr{C}$*) if there exists a prior π such that δ is ε-Bayes under π among $\mathscr{C}$ (resp., Bayes under π among $\mathscr{C}$).

We will sometimes write *Bayes among $\mathscr{C}$ with respect to π_0* to mean Bayes under π_0 among $\mathscr{C}$, and similarly for ε-Bayes among $\mathscr{C}$.

The following well-known result establishes a basic connection between Bayes optimality and admissibility (see, e.g., [8, Thm. 5.5.1]). We give a proof for completeness.

Theorem 6.1.8. *If δ is Bayes among $\mathscr{C}$, then δ is extended Bayes among $\mathscr{C}$, and then δ is extended admissible among $\mathscr{C}$.*

Proof. That Bayes implies extended Bayes follows trivially from definitions. Now assume δ is not extended admissible among $\mathscr{C}$. Then there exists $\varepsilon>0$ and $\delta'\in\mathscr{C}$ such that $r(\theta,\delta')\le r(\theta,\delta)-\varepsilon$ for all $\theta\in\Theta$. But then, for every prior π, $\int r(\theta,\delta')\pi(\mathrm{d}\theta)\le\int r(\theta,\delta)\pi(\mathrm{d}\theta)-\varepsilon$ or $\int r(\theta,\delta')\pi(\mathrm{d}\theta)=\int r(\theta,\delta)\pi(\mathrm{d}\theta)=\infty$, hence δ is not $\varepsilon/2$-Bayes among $\mathscr{C}$, hence not extended Bayes among $\mathscr{C}$. □

Note that neither extended admissibility nor admissibility imply Bayes optimality, in general. E.g., the maximum likelihood estimator in a univariate normal-location problem is admissible, but not Bayes.

Essential completeness allows us to strengthen a Bayes optimality claim:

Theorem 6.1.9. *Suppose $\mathscr{A}$ is an essentially complete subclass of $\mathscr{C}$, then ε-Bayes among $\mathscr{A}$ implies ε-Bayes among $\mathscr{C}$ for every $\varepsilon\ge 0$.*

Proof. Let δ_0 be Bayes under π among $\mathscr{A}$ for some prior π. Let $\delta\in\mathscr{C}$. Then there exists $\delta'\in\mathscr{A}$ such that, for all $r(\theta,\delta')\le r(\theta,\delta)$ for all $\theta\in\Theta$. By hypothesis, $r(\pi,\delta_0)\le r(\pi,\delta')$, but $r(\pi,\delta')=\int r(\theta,\delta')\pi(\mathrm{d}\theta)\le\int r(\theta,\delta)\pi(\mathrm{d}\theta)=r(\pi,\delta)$. Hence $r(\pi,\delta_0)\le r(\pi,\delta)$ for all $\delta\in\mathscr{C}$. □

6.1.3 Convexity

An important class of statistical decision problems are those in which the action space $\mathbb{A}$ is itself a vector space over the field $\mathbb{R}$. In that case, the mean estimate $\int_{\mathbb{A}} a\,\delta(x,\mathrm{d}a)$ is well defined for every $\delta \in \mathscr{D}_{0,FC}$ and $x \in X$, which motivates the following definition.

Definition 6.1.10. For $\delta \in \mathscr{D}_{0,FC}$, define $\mathbb{E}(\delta) : X \to \mathscr{M}_1(\mathbb{A})$ by $\mathbb{E}(\delta)(x,A) = 1$ if $\int_{\mathbb{A}} a\,\delta(x,\mathrm{d}a) \in A$ and 0 otherwise, for every $x \in X$ and measurable subset $A \subseteq \mathbb{A}$.

When the loss function is assumed to be convex, it is well known that the mean action will be no worse on average than the original randomized one. We formalize this condition below and prove several well-known results for completeness.

Condition LC (loss convexity)**.** $\mathbb{A}$ is a vector space over the field $\mathbb{R}$ and the loss function ℓ is convex with respect to the second argument.

Lemma 6.1.11. *Let δ and $\mathbb{E}(\delta)$ be as in Definition 6.1.10, and suppose (LC) holds. Then $r(\cdot,\delta) \geq r(\cdot,\mathbb{E}(\delta))$, hence $\mathbb{E}(\delta) \in \mathscr{D}_0$.*

Proof. Let $\theta \in \Theta$. By convexity of ℓ in its second parameter and a finite-dimensional version of Jensen's inequality [14, §2.8 Lem. 1], we have

$$r(\theta,\delta) = \int_X \Big[\int_{\mathbb{A}} \ell(\theta,a)\delta(x,\mathrm{d}a)\Big] P_\theta(\mathrm{d}x) \tag{6.1.5}$$

$$\geq \int_X \ell(\theta, \textstyle\int_{\mathbb{A}} a\,\delta(x,\mathrm{d}a)) P_\theta(\mathrm{d}x) = r(\theta,\mathbb{E}(\delta)). \tag{6.1.6}$$

□

Remark 6.1.12. Irrespective of the dimensionality of the action space $\mathbb{A}$, we may use a finite-dimensional version of Jensen's inequality because the procedure $\delta \in \mathscr{D}_{0,FC}$ is a finite mixture of nonrandomized procedures. The proof for a general randomized procedure $\delta \in \mathscr{D}$ and a general action space $\mathbb{A}$, would require additional hypotheses to account for the possible failure of Jensen's inequality (see [35]) and the possible lack of measurability of $\mathbb{E}(\delta)$ (see [14, S2.8]).

Lemma 6.1.13. *Suppose (LC) holds. Then $\mathscr{D}_0$ is an essentially complete subclass of $\mathscr{D}_{0,FC}$.*

Proof. Let $\delta \in \mathscr{D}_{0,FC}$. Then $\mathbb{E}(\delta) \in \mathscr{D}_0$. By Lemma 6.1.11, $\mathbb{E}(\delta)$ is well defined and $r(\theta,\delta_0) \geq r(\theta,\mathbb{E}(\delta))$, completing the proof. □

Remark 6.1.14. See the remark following [14, §2.8 Thm. 1] for a discussion of additional hypotheses needed for establishing that $\mathscr{D}_0$ is an essentially complete subclass of $\mathscr{D}$.

6.2 Prior Work

The first key results on admissibility and Bayes optimality are due to Abraham Wald, who laid the foundation of sequential decision theory. In [54], working in the setting of sequential statistical decision problems with compact parameter spaces, Wald showed that the Bayes decision procedures form an essentially complete class. Sequential decision problems differ from the decision problems we will be discussing in this paper in the sense that it gives the statistician the freedom to look at a sequence of observations one at a time and to decide, after each observation, whether to stop and take an action or to continue, potentially at some cost. The decision problems we will be discussing in this paper can be seen as special cases of sequential decision problems with only one observation.

In order to prove his results, Wald required a strong form of continuity for his risk and loss functions.

Definition 6.2.1. A sequence of parameters $\{\theta_i\}_{i\in\mathbb{N}}$ converges *in risk* to a parameter θ when $\sup_{\delta\in\mathscr{D}}|r(\theta_i,\delta)-r(\theta,\delta)|\to 0$ as $i\to\infty$, and converges *in loss* when $\sup_{a\in\mathbb{A}}|\ell(\theta_i,a)-\ell(\theta,a)|\to 0$ as $i\to\infty$. Similarly, a sequence of decision procedures $\{\delta_i\}_{i\in\mathbb{N}}$ in $\mathscr{D}$ converges *in risk* to a decision procedure δ when $\sup_{\theta\in\Theta}|r(\theta,\delta_i)-r(\theta,\delta)|\to 0$ as $i\to\infty$. A sequence of actions $\{a_i\}_{i\in\mathbb{N}}$ converges *in loss* to an action $a\in\mathbb{A}$ when $\sup_{\theta\in\Theta}|\ell(\theta,a_i)-\ell(\theta,a)|\to 0$ as $i\to\infty$.

Topologies on Θ, $\mathbb{A}$, and $\mathscr{D}$ are generated by these notions of convergence. In the following result and elsewhere, a model P is said to admit (a measurable family of) densities $(f_\theta)_{\theta\in\Theta}$ (with respect to a dominating (σ-finite) measure ν) when $P_\theta(A)=\int_A f_\theta(x)\,\nu(\mathrm{d}x)$ for every $\theta\in\Theta$ and measurable $A\subseteq X$. In terms of these densities, there is a unique Bayes solution with respect to a prior π on Θ when, for every $x\in X$, except perhaps for a set of ν-measure 0, there exists one and only one action $a^*\in\mathbb{A}$ for which the expression

$$\int_\Theta \ell(\theta,a)f_\theta(x)\,\pi(\mathrm{d}\theta) \tag{6.2.1}$$

takes its minimum value with respect to $a\in\mathbb{A}$. (Another notion of uniqueness used in the literature is to simply demand that the risk functions of two Bayes solutions agree.) The main result can be stated in the special case of a non-sequential decision problem as follows:

Theorem 6.2.2 ([54, Thms. 4.11 and 4.14]). *Assume Θ and $\mathscr{D}$ are compact in risk, and that Θ and $\mathbb{A}$ are compact in loss. Assume further that P admits densities $(f_\theta)_{\theta\in\Theta}$ with respect to Lebesgue measure, that these densities are strictly positive outside a Lebesgue measure zero set. Then every extended admissible decision procedure is Bayes. If the Bayes solution for every prior π is unique, the class of nonrandomized Bayes procedures form a complete class.*

Wald's regularity conditions are quite strong; he essentially requires equicontinuity in each variable for both the loss and risk functions. For example, the standard normal-location problem under squared error does not satisfy these criteria.

A similar result is established in the non-sequential setting in [55]:

Theorem 6.2.3 ([55, Thm. 3.1]). *Suppose that P admits densities $(f_\theta)_{\theta\in\Theta}$, that Θ is a compact subset of a Euclidean space, that the map $(x,\theta)\mapsto f_\theta(x)$ is jointly continuous, that the loss $\ell(\theta,a)$ is a continuous function of θ for every action a, that the space $\mathbb{A}$ is compact in loss, and that there is a unique Bayes solution for every prior π on Θ. Then every Bayes procedure is admissible and the collection of Bayes procedures form an essentially complete class.*

In many classical statistical decision problems, one does not lose anything by assuming that all risk functions are continuous. The following theorem, taken from [26], formalizes this intuition: We will say that a model P has a continuous likelihood function $(f_\theta)_{\theta\in\Theta}$ when P admits densities $(f_\theta)_{\theta\in\Theta}$ such that $\theta\mapsto f_\theta(x)$ is continuous for every $x\in X$.

Theorem 6.2.4 ([26, §5 Thm. 7.11]). *Suppose P has a continuous likelihood function $(f_\theta)_{\theta\in\Theta}$ and a monotone likelihood ratio. If the loss function $\ell(\theta,\delta)$ satisfies*

1. *$\ell(\theta,a)$ is continuous in θ for each action a;*
2. *$\ell(\theta,a)$ is decreasing in a for $a<\theta$ and increasing in a for $a>\theta$; and*
3. *there exist functions f and g, which are bounded on all bounded subsets of $\Theta\times\Theta$, such that for all a*

$$\ell(\theta,a)\le f(\theta,\theta')\ell(\theta',a)+g(\theta,\theta'), \tag{6.2.2}$$

then the estimators with finite-valued, continuous risk functions form a complete class.

If we assume the loss function is bounded, then all decision procedures have finite risk. The following theorem gives a characterization of continuous risk assuming boundedness of the loss.

Theorem 6.2.5 ([14, §3.7 Thm. 1])**.** *Suppose P admits densities $(f_\theta)_{\theta\in\Theta}$ with respect to a dominating measure ν. Assume*

1. *ℓ is bounded;*
2. *$\ell(\theta,a)$ is continuous in θ, uniformly in a;*
3. *for every bounded measurable ϕ, $\int \phi(x) f_\theta(x)\nu(\mathrm{d}x)$ is continuous in θ.*

Then the risk $r(\theta,\delta)$ is continuous in θ for every δ.

If we assume continuity of the risk function with respect to the parameter and restrict ourselves to Euclidean parameter spaces, we have the following theorem from [4, Sec. 8.8, Thm. 12].

Theorem 6.2.6. *Assume that $\mathbb{A}$ and Θ are compact subsets of Euclidean spaces and that the model P admits densities $(f_\theta)_{\theta\in\Theta}$ with respect to either Lebesgue or counting measure such that the map $(x,\theta)\mapsto f_\theta(x)$ is jointly continuous. Assume further that the loss $\ell(\theta,a)$ is a continuous function of $a\in\mathbb{A}$ for each θ, and that all decision procedures have continuous risk functions. Then the collection of Bayes procedures form a complete class.*

In the non-compact setting, Bayes procedures generally do not form a complete class. With a view to generalizing the notion of a Bayes procedure and recovering a complete class, Wald [56] introduced the notion of "Bayes in the wide sense", which we now call extended Bayes (see Definition 6.1.7). The formal statement of the following theorem is adapted from [14]:

Theorem 6.2.7. *Suppose that there exists a topology on $\mathscr{D}$ such that $\mathscr{D}$ is compact and $r(\theta,\delta)$ is lower semicontinuous in $\delta\in\mathscr{D}$ for all $\theta\in\Theta$. Then the set of extended Bayes procedures form an essentially complete class.*

Wald also studied taking the "closure" (in a suitable sense) of the collection of all Bayes procedures, and showed that every admissible procedure was contained in this new class. The first result of this form appears in [56] and is extended later in [25]. Brown [10, App. 4A] extended these results and gave a modern treatment. The following statement of Brown's version is adapted from [26, §5 Thm. 7.15].

Theorem 6.2.8. *Assume P admits strictly positive densities $(f_\theta)_{\theta\in\Theta}$ with respect to a σ-finite measure ν. Assume the action space $\mathbb{A}$ is a closed convex subset of Euclidean space. Assume the loss $\ell(\theta,a)$ is lower semicontinuous and strictly convex in a for every θ, and satisfies*

$$\lim_{|a|\to\infty} \ell(\theta,a) = \infty \text{ for all } \theta\in\Theta. \tag{6.2.3}$$

Then every admissible decision procedure δ is an a.e. limit of Bayes procedures, i.e., there exists a sequence π_n of priors with support on a finite set, such that

$$\delta^{\pi_n}(x) \to \delta(x) \text{ as } n \to \infty \text{ for } \nu\text{-almost all } x, \tag{6.2.4}$$

where δ^{π_n} is a Bayes procedure with respect to π_n.

In the normal-location model under squared error loss, the sample mean, while not a Bayes estimator in the strict sense, can be seen as a limit of Bayes estimators, e.g., with respect to normal priors of variance K as $K \to \infty$ or uniform priors on $[-K, K]$ as $K \to \infty$. (We revisit this problem in Example 8.3.2.) In his seminal paper, Sacks [43] observes that the sample mean is also the Bayes solution if the notion of prior distribution is relaxed to include Lebesgue measure on the real line. Sacks [43] raised the natural question: if δ is a limit of Bayes estimators, is there a measure m on the real line such that δ is "Bayes" with respect to this measure? A solution in this latter form was termed a *generalized Bayes solution* by Sacks [43]. The following definition is adapted from [52]:

Definition 6.2.9. A decision procedure δ_0 is a *normal-form generalized Bayes procedure* with respect to a σ-finite measure π on Θ when δ_m minimizes $r(\pi, \delta) = \int r(\theta, \delta)\pi(\mathrm{d}\theta)$, subject to the restriction that $r(\pi, \delta_m) < \infty$. If P admits densities $(f_\theta)_{\theta\in\Theta}$ with respect to a σ-finite measure ν and δ_0 minimizes the unnormalized posterior risk $\int \ell(\theta, \delta_0(x))\, f_\theta(x)\, \pi(\mathrm{d}\theta)$ for ν-a.e. x, then δ_0 is a *(extensive-form) generalized Bayes procedure* with respect to π.

When a model admits densities, Stone [52] showed that every normal-form generalized Bayes procedure is also extensive-form. (Sacks defined generalized Bayes in extensive form, but demanded also that $\int f_\theta(\cdot)\, \pi(\mathrm{d}\theta)$ be finite ν-a.e. The notion of normal- and extensive-form definitions of Bayes optimality were introduced by Raiffa and Schlaifer [37].) For exponential families, under suitable conditions, one can show that every admissible estimator is generalized Bayes. The first such result was developed by Sacks [43] in his original paper: he proved that, for statistical decision problems where the model admits a density of the form $e^{x\theta}/Z_\theta$ with $Z_\theta = \int e^{x\theta}\nu(\mathrm{d}\theta)$, every admissible estimator is generalized Bayes. Stone [52] extended this result to estimation of the mean in one-dimensional exponential families under squared error loss. These results were further generalized in similar ways by Brown [9, Sec. 3.1] and Berger and Srinivasan [5]. The following theorem is given in [5]. We adapt the statement of this theorem from [26].

Theorem 6.2.10 ([26, §5 Thm. 7.17])**.** *Assume the model is a finite-dimensional exponential family, and that the loss $\ell(\theta, a)$ is jointly continuous, strictly convex in a for every θ, and satisfies*

$$\lim_{|a| \to \infty} \ell(\theta, a) = \infty \text{ for all } \theta \in \Theta. \tag{6.2.5}$$

Then every admissible estimator is generalized Bayes.

Other generalized notions of Bayes procedures have been proposed. Heath and Sudderth [16] study statistical decision problems in the setting of finitely additive probability spaces. The following theorem is their main result:

Theorem 6.2.11 ([16, Thm. 2])**.** *Fix a class $\mathscr{D}$ of decision procedures. Every finitely additive Bayes decision procedure is extended admissible. If the loss function is bounded and the class $\mathscr{D}$ is convex, then every extended admissible decision procedure in $\mathscr{D}$ is finitely additive Bayes in $\mathscr{D}$.*

The simplicity of this statement is remarkable. However, the assumption of boundedness is very strong, and rule out many standard estimation problems on unbounded spaces. We will succeed in removing the boundedness assumption by moving to a sufficiently saturated nonstandard model.

Chapter 7

Nonstandard Statistical Decision Theory

As the literature stands, for infinite parameter spaces, the connection between frequentist and Bayesian optimality is subject to technical conditions, and these technical conditions (see Section 6.2) often rule out semi-parametric problems and regularly rule out nonparametric problems. As a result, the relationship between frequentist and Bayesian optimality in the setting of many modern statistical problems is uncharacterized. Indeed, given the effort expended to derive general results, it would be reasonable to assume that the connection between frequentist and Bayesian optimality was to some extent fragile, and might, in general, fail in nonparametric settings.

Using results in mathematical logic and nonstandard analysis, we identify an equivalence between the frequentist notion of extended admissibility (a necessary condition for both admissibility and minimaxity) and a novel notion of Bayesian optimality, and we show that this equivalence holds in *arbitrary* decision problems *without* technical conditions: informally, we show that, among decision procedures with finite risk functions, a decision procedure δ is extended admissible if and only if it has infinitesimal excess Bayes risk.

The fact that an equivalence holds, not just under weaker hypotheses than those employed in classical results, but under no assumptions, is surprising and suggests that our approach may be able to reveal further connections between frequentist and Bayesian optimality.

In Section 7.1, we define nonstandard counterparts of admissibility, extended admissibility, and essential completeness, which we obtain by ignoring infinitesimal violations of the standard notions, and then give key theorems relating standard and nonstandard notions for standard decision procedures and their nonstandard extensions, respectively.

In Section 7.2, we define a notion called nonstandard Bayes. Nonstandard Bayes is the nonstandard counterpart to Bayes optimality, which we also obtain by ignoring infinitesimal violations of the standard notion. We establish the connection between nonstandard Bayes and various notions of standard Bayes (Bayes, extended Bayes, generalized Bayes, etc). Using saturation and a hyperfinite version of the classical separating hyperplane argument on a hyperfinite discretization of the risk set, we show that a decision procedure is extended admissible if and only if it its nonstandard extension is nonstandard Bayes.

7.1 Nonstandard Admissibility

As we have seen in the previous section, strong regularity appears to be necessary to align Bayes optimality and admissibility. In non-compact parameter spaces, the statistician must apparently abandon the strict use of probability measures in order to represent certain extreme states of uncertainty that correspond with admissible procedures. Even then, strong regularity conditions are required (such as domination of the model and strict positiveness of densities, ruling out estimation in infinite-dimensional contexts). In the remainder of the paper, we describe a new approach using nonstandard analysis, in which the statistician uses probability measures, but has access to a much richer collection of real numbers with which to express their beliefs.

Let $(\Theta,\mathbb{A},\ell,X,P)$ be a standard statistical decision problem.

The nonstandard notions are the same as in previous chapters. For convenience of readers, we summarize them below. For a set S, let $\mathscr{P}(S)$ denote its power set. We assume that we are working within a nonstandard model containing $V \supseteq \mathbb{R}\cup\Theta\cup\mathbb{A}\cup X$, $\mathscr{P}(V), \mathscr{P}(V\cup\mathscr{P}(V)), \dots$, and we assume the model is as saturated as necessary. We use * to denote the nonstandard extension map taking elements, sets, functions, relations, etc., to their nonstandard counterparts. In particular, $^*\mathbb{R}$ and $^*\mathbb{N}$ denote the nonstandard extensions of the reals and natural numbers, respectively. Given a topological space (Y,T) and a subset $X \subseteq {}^*Y$, let $\mathrm{NS}(X) \subseteq X$ denote the subset of near-standard elements (defined by the monadic structure induced by T) and let $\mathsf{st} : \mathrm{NS}(Y) \to Y$ denote the standard part map taking near-standard elements to their standard parts. In both cases, the notation elides the underlying space Y and the topology T, because the space and topology will always be clear from context. As an abbreviation, we will write $^\circ x$ for $\mathsf{st}(x)$ for atomic elements x. For functions f, we will write $^\circ f$ for the composition $x \mapsto \mathsf{st}(f(x))$. Finally, given an internal (hyperfinitely additive) probability space $(\Omega,\mathscr{F},P)$, we will write $(\Omega,\overline{\mathscr{F}},\overline{P})$ to denote the corresponding Loeb space, i.e., the completion of the unique extension of P to $\sigma(\mathscr{F})$.

7.1.1 Nonstandard Extension of a Statistical Decision Problem

We will assume that Θ is a Hausdorff space and adopt its Borel σ-algebra $\mathscr{B}[\Theta]$.[1]

One should view the model P as a function from Θ to the space $\mathscr{M}_1(X)$ of probability measures on X. Write ${}^*P_y$ for $({}^*P)_y$. For every $y \in {}^*\Theta$, the transfer principle implies that ${}^*P_y$ is an internal probability measure on ${}^*X$ (defined on the extension of its σ-algebra). By the transfer principle, we know that ${}^*(P_\theta) = {}^*P_\theta$ for $\theta \in \Theta$, as one would expect from the notation.

Recall that standard decision procedures $\delta \in \mathscr{D}$ have finite risk functions. Therefore, the risk map $(\theta,\delta) \mapsto r(\theta,\delta)$ is a function from $\Theta \times \mathscr{D}$ to $\mathbb{R}$. By the extension and transfer principles, the nonstandard extension ${}^*r$ is an internal function from ${}^*\Theta \times {}^*\mathscr{D}$ to ${}^*\mathbb{R}$. and ${}^*\delta \in {}^*\mathscr{D}$ if $\delta \in \mathscr{D}$. The transfer principle also implies that every $\Delta \in {}^*\mathscr{D}$ is an internal function from ${}^*X$ to ${}^*\mathscr{M}_1(\mathbb{A})$. The *risk function of $\Delta \in {}^*\mathscr{D}$ is the function ${}^*r(\cdot,\Delta)$ from ${}^*\Theta$ to ${}^*\mathbb{R}$. By the transfer of the equation defining risk, the following statement holds:

$$(\forall \theta \in {}^*\Theta)\ (\forall \Delta \in {}^*\mathscr{D})\ ({}^*r(\theta,\Delta) = {}^*\!\!\int_{{}^*X} \left[{}^*\!\!\int_{{}^*\mathbb{A}} {}^*\ell(\theta,a)\Delta(x,\mathrm{d}a)\right] {}^*P_\theta(\mathrm{d}x). \tag{7.1.1}$$

As is customary, we will simply write $\int$ for ${}^*\!\int$, provided the context is clear. (We will also drop * from the extensions of common functions and relations like addition, multiplication, less-than-or-equal-to, etc.)

7.1.2 Nonstandard Admissibility

Let $\delta_0, \delta \in \mathscr{D}$, let $\varepsilon \in \mathbb{R}_{\geq 0}$, and assume δ_0 is ε-dominated by δ. Then there exists $\theta_0 \in \Theta$ such that

$$(\forall \theta \in \Theta)(r(\theta,\delta) \leq r(\theta,\delta_0) - \varepsilon) \wedge (r(\theta_0,\delta) \neq r(\theta_0,\delta_0)). \tag{7.1.2}$$

By the transfer principle,

$$(\forall \theta \in {}^*\Theta)({}^*r(\theta,{}^*\delta) \leq {}^*r(\theta,{}^*\delta_0) - \varepsilon) \wedge ({}^*r(\theta_0,{}^*\delta) \neq {}^*r(\theta_0,{}^*\delta_0)). \tag{7.1.3}$$

Because ${}^*r(\theta_0,{}^*\delta) = r(\theta_0,\delta)$ and similarly for ${}^*r(\theta_0,{}^*\delta_0)$, we know that ${}^*r(\theta_0,{}^*\delta) \not\approx {}^*r(\theta_0,{}^*\delta_0)$. These results motivate the following nonstandard version of domination.

[1] In one sense, this is a mild assumption, which we use to ensure that the standard part map $\mathsf{st} : \mathrm{NS}({}^*\Theta) \to \Theta$ is well defined. In another sense, Θ can always be made Hausdorff by, e.g., adopting the discrete topology. The topology determines the Borel sets and thus determines the set of available probability measures on Θ (and on ${}^*\Theta$, by extension). Topological considerations arise again in Section 8.1, Remark 8.2.8, and Remark 8.3.3.

Definition 7.1.1. Let $\Delta, \Delta' \in {}^*\mathscr{D}$ be internal decision procedures, let $\varepsilon \in \mathbb{R}_{\geq 0}$, and $R, S \subseteq {}^*\Theta$. Then Δ is ε-**dominated in R/S* by Δ' when

1. $\forall \theta \in S \;\; {}^*r(\theta, \Delta') \leq {}^*r(\theta, \Delta) - \varepsilon$, and
2. $\exists \theta \in R \;\; {}^*r(\theta, \Delta') \not\approx {}^*r(\theta, \Delta)$.

Write **dominated in R/S* for 0-*dominated in R/S, and write ε-**dominated on S* for ε-**dominated in S/S*.

The following results are immediate upon inspection of the definition above, and the fact that (1) implies (2) for $R \subseteq S$ when $\varepsilon > 0$.

Lemma 7.1.2. *Let $\varepsilon \leq \varepsilon'$, $R \subseteq R'$, and $S \subseteq S'$. Then ε'-*dominated in R/S' implies ε-*dominated in R'/S. If $\varepsilon > 0$, then ε-*dominated in S/S' if and only if ε-*dominated on S', and ε'-*dominated on S' implies ε-*dominated on S.*

The following result connects standard and nonstandard domination.

Theorem 7.1.3. *Let $\varepsilon \in \mathbb{R}_{\geq 0}$ and $\delta_0, \delta \in \mathscr{D}$. The following statements are equivalent:*

1. *δ_0 is ε-dominated by δ.*
2. *${}^*\delta_0$ is ε-*dominated in $\Theta/{}^*\Theta$ by ${}^*\delta$.*
3. *${}^*\delta_0$ is ε-*dominated on Θ by ${}^*\delta$.*

If $\varepsilon > 0$, then the following statement is also equivalent:

4. *${}^*\delta_0$ is ε-*dominated on ${}^*\Theta$ by ${}^*\delta$.*

Proof. ($1 \implies 2$) Follows from logic above Definition 7.1.1. ($2 \implies 3$) Follows from Lemma 7.1.2. ($3 \implies 1$) By hypothesis,

$$(\forall \theta \in \Theta)({}^*r(\theta, {}^*\delta) \leq {}^*r(\theta, {}^*\delta_0) - \varepsilon) \wedge (\exists \theta_0 \in \Theta)({}^*r(\theta_0, {}^*\delta) \not\approx {}^*r(\theta_0, {}^*\delta_0)). \tag{7.1.4}$$

Because ${}^*r(\theta_0, {}^*\delta) = r(\theta_0, \delta)$, and likewise for δ_0, it follows that

$$(\forall \theta \in \Theta)(r(\theta, \delta) \leq r(\theta, \delta_0) - \varepsilon). \tag{7.1.5}$$

Similarly, $^\circ({}^*r(\theta_0, {}^*\delta)) = r(\theta_0, \delta)$, and likewise for δ_0, hence **??**.1 implies

$$(\exists \theta_0 \in \Theta)(r(\theta_0, \delta) \neq r(\theta_0, \delta_0)). \tag{7.1.6}$$

(2 $\implies$ 4 $\implies$ 3) Follow from Lemma 7.1.2. □

Definition 7.1.4. Let $\varepsilon \in \mathbb{R}_{\geq 0}$, $R, S \subseteq {}^*\Theta$, and $\mathscr{C} \subseteq {}^*\mathscr{D}$, and $\Delta \in {}^*\mathscr{D}$.

1. Δ is ε-**admissible in R/S among $\mathscr{C}$* unless Δ is ε-*dominated in R/S by some $\Delta' \in \mathscr{C}$.
2. Δ is **admissible in R/S among $\mathscr{C}$* if Δ is 0-*admissible in R/S among $\mathscr{C}$.
3. Δ is ε-**admissible on S among $\mathscr{C}$* if Δ is ε-*admissible in S/S among $\mathscr{C}$.
4. Δ is **extended admissible on S among $\mathscr{C}$* if Δ is ε-*admissible on S among $\mathscr{C}$ for every $\varepsilon \in \mathbb{R}_{>0}$.

The following result is immediate upon inspection of the definitions above.

Lemma 7.1.5. *Let $\varepsilon \leq \varepsilon'$, $R \subseteq R'$, $S \subseteq S'$, and $\mathscr{A} \subseteq \mathscr{C}$. Then ε-*admissible in R'/S among $\mathscr{C}$ implies ε'-*admissible in R/S' among $\mathscr{A}$. For $\varepsilon > 0$, ε-*admissible on S among $\mathscr{C}$ implies ε'-*admissible on S' among $\mathscr{A}$.*

The analogous results for *admissible in R/S among $\mathscr{C}$ and *extended admissible on S among $\mathscr{C}$ then follow immediately. The following result connects standard and nonstandard admissibility. First, we must introduce the notion of the standard-part copy.

Definition 7.1.6. The standard-part copy of $\mathscr{C} \subseteq \mathscr{D}$ is ${}^\sigma\mathscr{C} = \{{}^*\delta : \delta \in \mathscr{C}\}$.

Note that ${}^\sigma\mathscr{C} \subseteq {}^*\mathscr{C}$ and ${}^\sigma\mathscr{C}$ is an external set unless $\mathscr{C}$ is finite.

Theorem 7.1.7. *Let $\varepsilon \in \mathbb{R}_{\geq 0}$, $\delta_0 \in \mathscr{D}$, and $\mathscr{C} \subseteq \mathscr{D}$. The following statements are equivalent:*

1. *δ_0 is ε-admissible among $\mathscr{C}$.*
2. *${}^*\delta_0$ is ε-*admissible in $\Theta/{}^*\Theta$ among ${}^\sigma\mathscr{C}$.*
3. *${}^*\delta_0$ is ε-*admissible on Θ among ${}^\sigma\mathscr{C}$.*

If $\varepsilon > 0$, then the following statements are also equivalent:

4. *${}^*\delta_0$ is ε-*admissible on ${}^*\Theta$ among ${}^\sigma\mathscr{C}$.*

5. ${}^*\delta_0$ *is* ε-**admissible on* ${}^*\Theta$ *among* ${}^*\mathscr{C}$.

Proof. Statement (1) is equivalent to

$$\neg(\exists \delta \in \mathscr{C})\ \delta_0 \text{ is } \varepsilon\text{-dominated by } \delta. \tag{7.1.7}$$

By Theorem 7.1.3 and the definition of ${}^\sigma\mathscr{C}$, this is equivalent to both

$$\neg(\exists^* \delta \in {}^\sigma\mathscr{C})\ {}^*\delta_0 \text{ is } \varepsilon\text{-}{}^*\text{dominated in } \Theta/{}^*\Theta \text{ by } {}^*\delta \tag{7.1.8}$$

and

$$\neg(\exists^* \delta \in {}^\sigma\mathscr{C})\ {}^*\delta_0\ \varepsilon\text{-}{}^*\text{dominated on } \Theta \text{ by } {}^*\delta, \tag{7.1.9}$$

hence $(1 \iff 2 \iff 3)$.

Now let $\varepsilon > 0$. Then the above statements are also equivalent to

$$\neg(\exists^* \delta \in {}^\sigma\mathscr{C})\ {}^*\delta_0 \text{ is } \varepsilon\text{-}{}^*\text{dominated on } {}^*\Theta \text{ by } {}^*\delta, \tag{7.1.10}$$

hence $(1 \iff 4)$. From Lemma 7.1.5, we see that (5) implies (4). To see that (1) implies (5), note that, because ε is standard and $\varepsilon > 0$, (1) is equivalent to

$$\neg(\exists \delta \in \mathscr{C})(\forall \theta \in \Theta)(r(\theta,\delta) \le r(\theta,\delta_0) - \varepsilon). \tag{7.1.11}$$

By transfer, this statement holds if and only if the following statement holds:

$$\neg(\exists \Delta \in {}^*\mathscr{C})(\forall \theta \in {}^*\Theta)({}^*r(\theta,\Delta) \le {}^*r(\theta,{}^*\delta_0) - \varepsilon). \tag{7.1.12}$$

Again, $\varepsilon > 0$ implies ${}^*r(\theta,\Delta) \not\approx {}^*r(\theta,{}^*\delta_0)$ for all $\theta \in {}^*\Theta$, hence (5) holds. □

The following corollary for extended admissibility follows immediately.

Theorem 7.1.8. *Let* $\delta_0 \in \mathscr{D}$ *and* $\mathscr{C} \subseteq \mathscr{D}$. *The following statements are equivalent:*

1. δ_0 *is extended admissible among* $\mathscr{C}$.

2. ${}^*\delta_0$ *is* **extended admissible on* Θ *among* ${}^\sigma\mathscr{C}$.

3. ${}^*\delta_0$ *is* **extended admissible on* ${}^*\Theta$ *among* ${}^\sigma\mathscr{C}$.

4. ${}^*\delta_0$ *is* **extended admissible on* ${}^*\Theta$ *among* ${}^*\mathscr{C}$.

As in the standard universe, the notion of *admissibility lead to notions of complete classes.

Definition 7.1.9. Let $\mathscr{A},\mathscr{C} \subseteq {}^*\mathscr{D}$.

1. $\mathscr{A}$ is a **complete subclass of* $\mathscr{C}$ if for all $\Delta \in \mathscr{C} \setminus \mathscr{A}$, there exists $\Delta' \in \mathscr{A}$ such that Δ is *dominated on Θ by Δ'.

2. $\mathscr{A}$ is an ${}^\circ$*essentially complete subclass of* $\mathscr{C}$ if for all $\Delta \in \mathscr{C} \setminus \mathscr{A}$, there exists $\Delta' \in \mathscr{A}$ such that ${}^*r(\theta,\Delta') \lessapprox {}^*r(\theta,\Delta)$ for all $\theta \in \Theta$.

Near-standard essential completeness allows us to enlarge the set of decision procedures amongst which a decision procedure is extended admissible.

Lemma 7.1.10. *Suppose* $\mathscr{A}$ *is an* ${}^\circ$*essentially complete subclass of* $\mathscr{C} \subseteq \mathscr{D}$. *Then* **extended admissible on* Θ *among* $\mathscr{A}$ *implies* **extended admissible on* Θ *among* $\mathscr{C}$.

Proof. Let $\Delta_0 \in \mathscr{A}$ and suppose Δ_0 is not *extended admissible on Θ among $\mathscr{C}$. Then there exists $\Delta \in \mathscr{C}$ and $\varepsilon \in \mathbb{R}_{>0}$ such that ${}^*r(\theta,\Delta) \leq {}^*r(\theta,\Delta_0) - \varepsilon$ for all $\theta \in \Theta$. But then by the *essential completeness of $\mathscr{A}$, there exists some $\Delta' \in \mathscr{A}$, such that ${}^*r(\theta,\Delta') \lessapprox {}^*r(\theta,\Delta)$ for all $\theta \in \Theta$, hence ${}^*r(\theta,\Delta') \lessapprox {}^*r(\theta,\Delta_0) - \varepsilon$ for all $\theta \in \Theta$. But then Δ_0 is not $\varepsilon/2$-*admissible on Θ among $\mathscr{A}$ hence not *extended admissible on Θ among $\mathscr{A}$. □

7.2 Nonstandard Bayes

We now define the nonstandard counterparts to Bayes risk and optimality for the class ${}^*\mathscr{D}$ of internal decision procedures:

Definition 7.2.1. Let $\Delta \in {}^*\mathscr{D}$, $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$, and $\mathscr{C} \subseteq {}^*\mathscr{D}$, and let Π_0 be a *nonstandard prior*, i.e., an internal probability measure on $({}^*\Theta, {}^*\mathscr{B}[\Theta])$. The *internal Bayes risk* under Π_0 of Δ is ${}^*r(\Pi_0,\Delta) = \int {}^*r(\theta,\Delta)\Pi_0(\mathrm{d}\theta)$.

1. Δ is ε-**Bayes under* Π_0 *among* $\mathscr{C}$ if ${}^*r(\Pi_0,\Delta)$ is hyperfinite and, for all $\Delta' \in \mathscr{C}$, we have ${}^*r(\Pi_0,\Delta) \leq {}^*r(\Pi_0,\Delta') + \varepsilon$.

2. Δ is *nonstandard Bayes under* Π_0 *among* $\mathscr{C}$ if ${}^*r(\Pi_0,\Delta)$ is hyperfinite and, for all $\Delta' \in \mathscr{C}$, we have ${}^*r(\Pi_0,\Delta) \lessapprox {}^*r(\Pi_0,\Delta')$. [2]

We will write *nonstandard Bayes among* $\mathscr{C}$ *with respect to* Π_0 to mean nonstandard Bayes under Π_0 among $\mathscr{C}$ and will write *nonstandard Bayes among* $\mathscr{C}$ to mean *nonstandard Bayes among* $\mathscr{C}$ with respect to some nonstandard prior Π. The same abbreviations will be used for ε-*Bayes among $\mathscr{C}$. Note that the internal Bayes risk is precisely the extension of the standard Bayes risk. Similarly, if we consider the relation $\{(\delta,\varepsilon,\mathscr{C}) \in \mathscr{D} \times \mathbb{R}_{\geq 0} \times \mathscr{P}(\mathscr{D}) : \delta \text{ is } \varepsilon\text{-Bayes among } \mathscr{C}\}$, then its extension corresponds to $\{(\Delta,\varepsilon,\mathscr{C}) \in {}^*\mathscr{D} \times {}^*\mathbb{R}_{\geq 0} \times {}^*\mathscr{P}(\mathscr{D}) : \Delta \text{ is } \varepsilon\text{-}{}^*\text{Bayes among } \mathscr{C}\}$. Note, however, that our definition of "ε-*Bayes among $\mathscr{C}$" allows the set $\mathscr{C} \subseteq {}^*\mathscr{D}$ to be external, and so it is not simply the transfer of the standard relation. The following lemma relates the two nonstandard notions of Bayes optimality: Recall that our nonstandard model is κ saturated.

Lemma 7.2.2. *Let* $\mathscr{C} \subseteq {}^*\mathscr{D}$. *If* $\varepsilon \approx 0$, *then* ε-**Bayes under* Π_0 *among* $\mathscr{C}$ *implies nonstandard Bayes under* Π_0 *among* $\mathscr{C}$. *In the other direction, if* $\mathscr{C}$ *is either internal or has a fixed external cardinality less than* κ, *then nonstandard Bayes under* Π_0 *among* $\mathscr{C}$ *implies* ε-**Bayes under* Π_0 *among* $\mathscr{C}$ *for some* $\varepsilon \approx 0$.

Proof. The first statement is trivial. Suppose Δ_0 is nonstandard Bayes under Π_0 among $\mathscr{C}$. By definition, we have ${}^*r(\Pi_0,\Delta_0) \lessapprox {}^*r(\Pi_0,\Delta)$ for all $\Delta \in \mathscr{C}$. Let

$$A = \{|{}^*r(\Pi_0,\Delta_0) - {}^*r(\Pi_0,\Delta)| : \Delta \in \mathscr{C}\} \tag{7.2.1}$$

and

$$A^n_\Delta = \{\varepsilon \in {}^*\mathbb{R} : |{}^*r(\Pi_0,\Delta_0) - {}^*r(\Pi_0,\Delta)| \leq \varepsilon \leq \frac{1}{n}\}. \tag{7.2.2}$$

If $\mathscr{C}$ is internal, then A is internal and so it has a least upper bound ε. Because A contains only infinitesimals, $\varepsilon \approx 0$, for otherwise $\varepsilon/2$ would also be an upper bound on A. Thus, we have ${}^*r(\Pi_0,\Delta_0) \leq {}^*r(\Pi_0,\Delta) + \varepsilon$ for all $\Delta \in \mathscr{C}$ which shows that Δ_0 is ε-*Bayes under Π_0 among $\mathscr{C}$.

If $\mathscr{C}$ has a fixed external cardinality less than κ, then $\mathscr{F} = \{A^n_\Delta : \Delta \in \mathscr{C}, n \in \mathbb{N}\}$ has a fixed external cardinality less than κ. It is easy to see that $\mathscr{F}$ has the finite intersection property. By saturation, the

[2] The definition of nonstandard Bayes is obtained by extending the standard definition of Bayes, but allowing for infinitesimal violations of the criterion. There was a consensus to denote such notion with the prefix "S-" rather than "nonstandard". However, we use "nonstandard Bayes" instead of "S-Bayes" to emphasize on the fact that this definition is nonstandard.

total intersection of $\mathscr{F}$ is non-empty. That is, there exists $\varepsilon_0 \in {}^*\mathbb{R}$ such that $\varepsilon_0 \leq \frac{1}{n}$ for all $n \in \mathbb{N}$ and $\varepsilon_0 \geq |{}^*r(\Pi_0, \Delta_0) - {}^*r(\Pi_0, \Delta)|$ for all $\Delta \in \mathscr{C}$. Thus $\varepsilon_0 \approx 0$ and Δ_0 is ε_0-*Bayes under Π_0 among $\mathscr{C}$. □

Transfer remains a powerful tool for relating the optimality of standard procedures with that of their extensions. For example, by transfer, δ is ε-Bayes under π among $\mathscr{C}$ if and only if ${}^*\delta$ is ε-*Bayes under ${}^*\pi$ among ${}^*\mathscr{C}$. (Recall that ${}^*\varepsilon = \varepsilon$ for a real ε, by extension.) Transfer also yields the following result:

Theorem 7.2.3. *Let $\delta_0 \in \mathscr{D}$ and $\mathscr{C} \subseteq \mathscr{D}$. The following statements are equivalent:*

1. *δ_0 is extended Bayes among $\mathscr{C}$.*
2. *${}^*\delta_0$ is ε-*Bayes among ${}^*\mathscr{C}$ for all $\varepsilon \in {}^*\mathbb{R}_{>0}$.*
3. *${}^*\delta_0$ is ε_0-*Bayes among ${}^*\mathscr{C}$ for some $\varepsilon_0 \approx 0$.*
4. *${}^*\delta_0$ is nonstandard Bayes among ${}^*\mathscr{C}$.*

Proof. Suppose δ_0 is extended Bayes among $\mathscr{C}$. By hypothesis, the following sentence holds:

$$(\forall \varepsilon \in \mathbb{R}_{>0})(\exists \pi \in \mathscr{M}_1(\Theta))(\forall \delta \in \mathscr{C})(r(\pi, \delta_0) \leq r(\pi, \delta) + \varepsilon). \tag{7.2.3}$$

By the transfer principle,

$$(\forall \varepsilon \in {}^*\mathbb{R}_{>0})(\exists \pi \in {}^*\mathscr{M}_1(\Theta))(\forall \delta \in {}^*\mathscr{C})({}^*r(\pi, {}^*\delta_0) \leq {}^*r(\pi, {}^*\delta) + \varepsilon). \tag{7.2.4}$$

Thus, (1) implies (2). It is clear that (2) implies (3). By Lemma 7.2.2, we know that (3) implies (4).

Now suppose ${}^*\delta_0$ is nonstandard Bayes among ${}^*\mathscr{C}$. Pick $\varepsilon \in \mathbb{R}_{>0}$. It is easy to see that ${}^*\delta_0$ is ε-*Bayes among ${}^*\mathscr{C}$. Formally,

$$(\exists \pi \in {}^*\mathscr{M}_1(\Theta))(\forall \delta \in {}^*\mathscr{C})({}^*r(\pi, {}^*\delta_0) \leq {}^*r(\pi, {}^*\delta) + \varepsilon). \tag{7.2.5}$$

By the transfer principle,

$$(\exists \pi \in \mathscr{M}_1(\Theta))(\forall \delta \in \mathscr{C})(r(\pi, \delta_0) \leq r(\pi, \delta) + \varepsilon). \tag{7.2.6}$$

Thus δ_0 is ε-Bayes among $\mathscr{C}$. As ε was chosen arbitrarily, δ_0 is extended Bayes among $\mathscr{C}$. □

We also establish the following result that connects normal-form generalized Bayes and nonstandard Bayes.

Theorem 7.2.4. *Let $\delta_0 \in \mathscr{D}$ be normal-form generalized Bayes among $\mathscr{C} \subset \mathscr{D}$. Then ${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$.*

Proof. Let μ be a nonzero σ-finite measure with respect to which δ_0 is normal-form generalized Bayes among $\mathscr{C}$. As μ is σ-finite, we can write $\Theta = \bigcup_{n\in\mathbb{N}} V_n$ where $V_i \subset V_j$ for $i \leq j$ and $\mu(V_n) \in \mathbb{R}_{>0}$ for all $n \in \mathbb{N}$. By extension, there exists an internal sequence of *measurable sets $\{U_n : n \in {}^*\mathbb{N}\}$ satisfying the following conditions:

- $U_n = {}^*V_n$, for $n \in \mathbb{N}$,
- $U_i \subset U_j$, for $i \leq j \in {}^*\mathbb{N}$, and
- ${}^*\mu(U_n) \in {}^*\mathbb{R}_{>0}$, for all $n \in \mathbb{N}$.

Let $\mathscr{F}(\mathscr{C}) = \{\delta \in \mathscr{C} : r(\mu, \delta) < \infty\}$ and fix an infinitesimal $\varepsilon > 0$. For every $\delta \in \mathscr{F}(\mathscr{C})$, the transfer principle implies there exists $N_\delta \in {}^*\mathbb{N}$ such that

$$\int_{U_{N_\delta}} {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) \geq \int {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) - \varepsilon. \tag{7.2.7}$$

Then, by saturation, there exists $N \in {}^*\mathbb{N}$ such that

$$\int_{U_k} {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) \geq \int {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) - \varepsilon \tag{7.2.8}$$

for all $\delta \in \mathscr{F}(\mathscr{C})$ and all $k \geq N$. By the generalized Bayes optimality of ${}^*\delta_0$ and the transfer principle, $\int {}^*r(\theta, {}^*\delta_0)\,{}^*\mu(\mathrm{d}\theta) \leq \int {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta)$ for all $\delta \in \mathscr{F}(\mathscr{C})$. As ε is infinitesimal, we have

$$\int_{U_k} {}^*r(\theta, {}^*\delta_0)\,{}^*\mu(\mathrm{d}\theta) \lessapprox \int_{U_k} {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) \tag{7.2.9}$$

for all $\delta \in \mathscr{F}(\mathscr{C})$ and all $k \geq N$.

By the saturation principle, there exists $r \in {}^*\mathbb{R}_{>0}$ such that $\int {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) < r$ for all $\delta \in \mathscr{F}(\mathscr{C})$. By transfer and then saturation, there exists $N' \in {}^*\mathbb{N}$ such that

$$\int_{U_{N'}} {}^*r(\theta, {}^*d)\,{}^*\mu(\mathrm{d}\theta) > r \tag{7.2.10}$$

for all $d \in \mathscr{C} \setminus \mathscr{F}(\mathscr{C})$. Let $N_0 = \max\{N, N'\}$. By Eqs. (7.2.9) and (7.2.10), we have

$$\int_{U_{N_0}} {}^*r(\theta, {}^*\delta_0)\,{}^*\mu(\mathrm{d}\theta) \lessapprox \int_{U_{N_0}} {}^*r(\theta, {}^*\delta)\,{}^*\mu(\mathrm{d}\theta) \tag{7.2.11}$$

for all $\delta \in \mathscr{C}$. Because ${}^*\mu(U_{N_0}) \in {}^*\mathbb{R}_{>0}$, the quantity $\pi(A) = \frac{{}^*\mu(A \cap U_{N_0})}{{}^*\mu(U_{N_0})}$ is well defined for every *measurable set $A \subseteq {}^*\Theta$. It is easy to see that π is an internal probability measure on ${}^*\Theta$. Moreover, by Eq. (7.2.11), ${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$ with respect to π. □

Remark 7.2.5. Note that our model is more saturated than the cardinality of $\mathscr{D}$, and so Lemma 7.2.2 implies that ${}^*\delta_0$ is even ε_0-*Bayes among ${}^\sigma\mathscr{C}$ for some $\varepsilon_0 \approx 0$.

Example 7.2.6. Consider the classical normal-location problem with squared error loss. It is well known that the maximum likelihood estimator $\delta(x) = x$ is normal-form generalized Bayes among all estimators with respect to the Lebesgue measure μ on $\mathbb{R}$. Inspecting the proof of Theorem 7.2.4, we see that there exists an infinite $K \in {}^*\mathbb{R}_{\geq 0}$ such that ${}^*\delta$ is nonstandard Bayes with respect to the internal uniform probability measure on $[-K, K]$.

In general, we would not expect the extension of a standard procedure to be 0-*Bayes under Π among $\mathscr{C}$ for a generic nonstandard prior Π and class $\mathscr{C} \subseteq {}^*\mathscr{D}$. The definition of nonstandard Bayes provides infinitesimal slack, which suffices to yield a precise characterization of extended admissible procedures. The following result shows that nonstandard Bayes optimality implies nonstandard extended admissibility, much like in the standard universe.

Theorem 7.2.7. *Let $\Delta_0 \in {}^*\mathscr{D}$, let $\mathscr{C} \subseteq {}^*\mathscr{D}$, and suppose that Δ_0 is nonstandard Bayes among $\mathscr{C}$. Then Δ_0 is *extended admissible on ${}^*\Theta$ among $\mathscr{C}$.*

Proof. Suppose Δ_0 is not *extended admissible on ${}^*\Theta$ among $\mathscr{C}$. Then for some standard $\varepsilon \in \mathbb{R}_{>0}$, Δ_0 is ε-*dominated on ${}^*\Theta$ by some $\Delta \in \mathscr{C}$, i.e.,

$$(\forall \theta \in {}^*\Theta)({}^*r(\theta, \Delta) \leq {}^*r(\theta, \Delta_0) - \varepsilon). \tag{7.2.12}$$

Hence, for every nonstandard prior Π, if ${}^*r(\Pi, \Delta)$ is not hyperfinite, then neither is ${}^*r(\Pi, \Delta_0)$, and if ${}^*r(\Pi, \Delta)$ is hyperfinite, then

$$\begin{aligned} {}^*r(\Pi, \Delta_0) &= \int {}^*r(\theta, \Delta_0)\Pi(\mathrm{d}\theta) & (7.2.13)\\ &\geq \int {}^*r(\theta, \Delta)\Pi(\mathrm{d}\theta) + \varepsilon = {}^*r(\Pi, \Delta) + \varepsilon. & (7.2.14) \end{aligned}$$

As $\varepsilon \in \mathbb{R}_{>0}$, we conclude that Δ_0 cannot be nonstandard Bayes under Π among $\mathscr{C}$. As Π was arbitrary, Δ_0 is not nonstandard Bayes among $\mathscr{C}$. □

Theorems 7.1.8 and 7.2.7 immediately yield the following corollary.

Corollary 7.2.8. *Let $\delta \in \mathscr{D}$ and $\mathscr{C} \subseteq \mathscr{D}$. If ${}^*\delta$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$, then δ is extended admissible among $\mathscr{C}$.*

The above result raises several questions: Are extended admissible decision procedures also nonstandard Bayes? What is the relationship with admissibility and its nonstandard counterparts?

In this section, we prove that a decision procedure δ is extended admissible if and only if ${}^*\delta$ is nonstandard Bayes. In later sections, we give several application of this equivalence, and then consider the relationship with admissibility, which is far from settled. It is easy, however, to show that only nonstandard Bayes procedures can *dominate other nonstandard Bayes procedures: To see this, suppose that Δ is nonstandard Bayes among $\mathscr{C} \subseteq {}^*\mathscr{D}$ with respect to some nonstandard prior Π and Δ is not *admissible on ${}^*\Theta$ among $\mathscr{C}$.

Then Δ is *dominated on ${}^*\Theta$ by some $\Delta' \in \mathscr{C}$. Thus we have ${}^*r(\theta,\Delta') \leq {}^*r(\theta,\Delta)$ for all $\theta \in {}^*\Theta$. By Definition 7.2.1, we have ${}^*r(\Pi,\Delta) = \int {}^*r(\theta,\Delta)\Pi(\mathrm{d}\theta)$ hyperfinite. But then, ${}^*r(\Pi,\Delta) \lessapprox {}^*r(\Pi,\Delta') = \int {}^*r(\theta,\Delta')\Pi(\mathrm{d}\theta) \leq {}^*r(\Pi,\Delta)$, hence ${}^*r(\Pi,\Delta) \approx {}^*r(\Pi,\Delta')$, hence Δ' is nonstandard Bayes under Π among $\mathscr{C}$. This proves a nonstandard version of a well-known standard result stating that every unique Bayes procedure is admissible [14, §2.3 Thm. 1]:

Theorem 7.2.9. *Suppose Δ is nonstandard Bayes among $\mathscr{C} \subseteq {}^*\mathscr{D}$ with respect to a nonstandard prior Π. If Δ is *dominated on ${}^*\Theta$ by $\Delta' \in \mathscr{C}$, then Δ' is nonstandard Bayes under Π among $\mathscr{C}$. Therefore, if ${}^*r(\theta,\Delta') \approx {}^*r(\theta,\Delta)$ for all $\theta \in {}^*\Theta$ and for all $\Delta' \in \mathscr{C}$ such that Δ' is nonstandard Bayes under Π among $\mathscr{C}$, then Δ is *admissible on ${}^*\Theta$ among $\mathscr{C}$.*

Proof. The first statement follows from the logic in the preceding paragraph. Now suppose that Δ is *dominated on ${}^*\Theta$ by some $\Delta' \in \mathscr{C}$. Then Δ' is nonstandard Bayes under Π among $\mathscr{C}$. But then, by hypothesis, its risk function is equivalent, up to an infinitesimal, to that of Δ, a contradiction. □

7.2.1 Hyperdiscretized Risk Set

In a statistical decision problem with a finite parameter space, one can use a separating hyperplane argument to show that every admissible decision procedure is Bayes (see, e.g., [14, §2.10 Thm. 1]). In order to prove our main theorem, we will proceed along similar lines, but with the aid of extension, transfer, and saturation.

When relating extended admissibility and Bayes optimality for a subclass $\mathscr{C} \subseteq \mathscr{D}$, the set of all risk functions r_δ, for $\delta \in \mathscr{C}$, is a key structure. On a finite parameter space, the risk set for $\mathscr{D}$ is a convex subset of a finite-dimensional vector space over $\mathbb{R}$. When the parameter space is not finite, one must grapple with infinite dimensional function spaces. However, in a sufficiently saturated nonstandard model, there exists an internal set $T_\Theta \subset {}^*\Theta$ that is hyperfinite and contains Θ. While the risk at all points in T_Θ does not suffice to characterize an arbitrary element of ${}^*\mathscr{D}$, it suffices to study the optimality of extensions of standard decision procedure *relative to other extensions*. Because T_Θ is hyperfinite, the corresponding risk set is a convex subset of a hyperfinite-dimensional vector space over ${}^*\mathbb{R}$.

Let $J_\Theta \in {}^*\mathbb{N}$ be the internal cardinality of T_Θ and let $T_\Theta = \{t_1, \dots, t_{J_\Theta}\}$. Recall that $\mathrm{I}({}^*\mathbb{R}^{J_\Theta})$ denotes the set of (internal) functions from T_Θ to ${}^*\mathbb{R}$. For an element $x \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})$, we will write x_k for $x(k)$.

Definition 7.2.10. The *hyperdiscretized risk set induced by $D \subseteq {}^*\mathscr{D}$* is the set

$$\mathscr{S}^D = \{x \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta}) : (\exists \Delta \in D)\,(\forall k \le J_\Theta)\, x_k = {}^*r(t_k, \Delta)\} \subset \mathrm{I}({}^*\mathbb{R}^{J_\Theta}). \tag{7.2.15}$$

Lemma 7.2.11. *Let $D \subseteq {}^*\mathscr{D}$ be an internal convex set. Then $\mathscr{S}^D$ is an internal convex set.*

Proof. $\mathscr{S}^D$ is internal by the internal definition principle and the fact that D is internal. In order to demonstrate convexity, pick $p \in {}^*[0,1]$, and let $x, y \in \mathscr{S}^D$. Then there exist $\Delta_1, \Delta_2 \in D$ such that $x_k = {}^*r(t_k, \Delta_1)$ and $y_k = {}^*r(t_k, \Delta_2)$ for all $k \le J_\Theta$. Because D is convex, $p\Delta_1 + (1-p)\Delta_2 \in D$. But $px_k + (1-p)y_k = {}^*r(t_k, p\Delta_1 + (1-p)\Delta_2)$ for all $k \le J_\Theta$, and so $\mathscr{S}^D$ is convex. ☐

Definition 7.2.12. For every $\mathscr{C} \subseteq {}^*\mathscr{D}$, let

$$(\mathscr{C})_{FC} = \bigcup_{D \in \mathscr{C}^{[<\infty]}} {}^*\mathrm{conv}(D) \tag{7.2.16}$$

be the set of all finite *convex combinations of ${}^*\delta \in \mathscr{C}$.

Let $\delta_1, \delta_2 \in \mathscr{D}_0$ and let $p \in {}^*[0,1]$. Then $p{}^*\delta_1 + (1-p){}^*\delta_2 \in {}^\sigma\mathscr{D}_{0,FC}$ if $p \in [0,1]$. However, $p{}^*\delta_1 + (1-p){}^*\delta_2 \in ({}^\sigma\mathscr{D}_0)_{FC}$ for all $p \in {}^*[0,1]$. It is easy to see that $({}^\sigma\mathscr{D}_{0,FC})_{FC} = ({}^\sigma\mathscr{D}_0)_{FC}$. Thus, we have ${}^\sigma\mathscr{D}_0 \subset {}^\sigma\mathscr{D}_{0,FC} \subset ({}^\sigma\mathscr{D}_{0,FC})_{FC} = ({}^\sigma\mathscr{D}_0)_{FC} \subset {}^*\mathscr{D}_{0,FC}$.

Lemma 7.2.13. *For any $\mathscr{C} \subseteq {}^*\mathscr{D}$, $(\mathscr{C})_{FC}$ is a convex set containing $\mathscr{C}$.*

Proof. Pick an $\mathscr{C} \subseteq {}^*\mathscr{D}$. Clearly $(\mathscr{C})_{FC} \supset \mathscr{C}$. It remains to show that $(\mathscr{C})_{FC}$ is a convex set. Pick two elements $\Delta_1, \Delta_2 \in (\mathscr{C})_{FC}$. Then there exist $D_1, D_2 \in \mathscr{C}^{[<\infty]}$ such that $\Delta_1 \in {}^*\mathrm{conv}(D_1)$ and $\Delta_2 \in {}^*\mathrm{conv}(D_2)$. Let $p \in {}^*[0,1]$. It is easy to see that $p\Delta_1 + (1-p)\Delta_2 \in {}^*\mathrm{conv}(D_1 \cup D_2)$. ☐

Lemma 7.2.14. *${}^\sigma\mathscr{D}_{0,FC}$ is an ${}^\circ$essentially complete subclass of $({}^\sigma\mathscr{D}_0)_{FC}$.*

Proof. Let $\Delta \in ({}^{\sigma}\mathscr{D}_0)_{FC}$. Then $\Delta = \sum_{i=1}^{n} p_i {}^*\delta_i$ for some $n \in \mathbb{N}$, $\delta_1, \dots, \delta_n \in \mathscr{D}_0$, and $p_1, \dots, p_n \in {}^*\mathbb{R}_{\geq 0}$, $\sum_{i=1}^{n} p_i = 1$. Define $\Delta_0 = \sum_{i=1}^{n} {}^{\circ}p_i {}^*\delta_i$ and let $\theta \in \Theta$. For all $i \leq n$, $p_i {}^*r(\theta, {}^*\delta_i) \approx {}^{\circ}p_i {}^*r(\theta, {}^*\delta_i)$ because ${}^*r(\theta, {}^*\delta_i)$ is finite, so ${}^*r(\theta, \Delta) \approx {}^*r(\theta, \Delta_0)$. By Definition 7.1.9, ${}^{\sigma}\mathscr{D}_{0,FC}$ is an ${}^{\circ}$essentially complete subclass of $({}^{\sigma}\mathscr{D}_0)_{FC}$. $\square$

Having defined the (hyperdiscretized) risk set, we now describe a set whose intersection with the risk set captures the notion of $\frac{1}{n}$-*domination, for some standard $n \in \mathbb{N}$. In that vein, for $\Delta \in {}^*\mathscr{D}$, define the $\frac{1}{n}$-*quantant*

$$Q(\Delta)_n = \{x \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta}) : (\forall k \leq J_\Theta)(x_k \leq {}^*r(t_k, \Delta) - \frac{1}{n})\}, \qquad n \in {}^*\mathbb{N}. \tag{7.2.17}$$

Lemma 7.2.15. *Fix $\Delta \in {}^*\mathscr{D}$. The set $Q(\Delta)_n$ is internal and convex and $Q(\Delta)_m \subset Q(\Delta)_n$ for every $m < n$.*

Proof. By the internal definition principle, $Q(\Delta)_n$ is internal. Let x, y be two points in $Q(\Delta)_n$, let $p \in {}^*[0,1]$, and pick a coordinate k. Then

$$px_k + (1-p)y_k \leq p({}^*r(t_k, \Delta) - \frac{1}{n}) + (1-p)({}^*r(t_k, \Delta) - \frac{1}{n}) = ({}^*r(t_k, \Delta) - \frac{1}{n}). \tag{7.2.18}$$

Thus the set is convex. The second statement is obvious. $\square$

The following is then immediate from definitions.

Lemma 7.2.16. *Let $\mathscr{C} \subseteq {}^*\mathscr{D}$ and $n \in \mathbb{N}$. Then Δ is $\frac{1}{n}$-*admissible on T_Θ among $\mathscr{C}$ if and only if $Q(\Delta)_n \cap \mathscr{S}^{\mathscr{C}} = \emptyset$.*

7.2.2 Nonstandard Complete Class Theorems

Lemma 7.2.17. *Let $\Delta \in {}^*\mathscr{D}$ and nonempty $D \subseteq {}^*\mathscr{D}$, and suppose there exists a nonzero vector $\Pi \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})$ such that $\langle \Pi, x \rangle \leq \langle \Pi, s \rangle$ for all $x \in \bigcup_{n \in \mathbb{N}} Q(\Delta)_n$ and $s \in \mathscr{S}^D$. Then the normalized vector $\Pi / \|\Pi\|_1$ induces an internal probability measure π on ${}^*\Theta$ concentrating on T_Θ, and Δ is nonstandard Bayes under π among D.*

Proof. We first establish that $\Pi(k) \geq 0$ for all k. Suppose otherwise, i.e., $\Pi(k_0) < 0$ for some k_0. Then we can pick a point x_0 in $\bigcup_{n \in \mathbb{N}} Q(\Delta)_n$ whose k_0-th coordinate is arbitrarily large and negative, causing $\langle \Pi, x_0 \rangle$ to be arbitrary large, a contradiction because $\langle \Pi, s \rangle$ is hyperfinite for all $s \in \mathscr{S}^D$. Hence, all coordinates of Π must be nonnegative.

Define $\pi \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})$ by $\pi = \Pi/\|\Pi\|_1$. Because $\Pi \neq 0$ and $\Pi \geq 0$, we have $\pi \geq 0$ and $\|\pi\|_1 = 1$. Therefore, π specifies an internal probability measure on $({}^*\Theta, {}^*\mathscr{B}[\Theta])$, concentrating on T_Θ, and assigning probability $\pi(k)$ to t_k for every $k \leq J_\Theta$. Because $\|\Pi\|_1 > 0$, it still holds that $\langle \pi, x\rangle \leq \langle \pi, s\rangle$ for all $x \in \bigcup_{n\in\mathbb{N}} Q(\Delta)_n$ and $s \in \mathscr{S}^D$.

Let $s \in \mathscr{S}^D$. Then ${}^\circ(\sum_{k\in J_\Theta} \pi_k({}^*r(t_k,\Delta) - \frac{1}{n})) \leq {}^\circ(\sum_{k\in J_\Theta} \pi_k s_k)$ for every $n \in \mathbb{N}$. The l.h.s. is simply ${}^\circ(-\frac{1}{n} + \sum_{k\in J_\Theta} \pi_k {}^*r(t_k,\Delta))$, and the limit of this expression as $n \to \infty$ is ${}^\circ(\sum_{k\in J_\Theta} \pi_k {}^*r(t_k,\Delta))$. Hence, $\sum_{k\in J_\Theta} \pi_k({}^*r(t_k,\Delta) \lessapprox \sum_{k\in J_\Theta} \pi_k s_k$. This shows that Δ is nonstandard Bayes under π among D. □

The previous result shows that if a nontrivial hyperplane separates the risk set from every $\frac{1}{n}$-quantant, for $n \in \mathbb{N}$, then the corresponding procedure is nonstandard Bayes. In order to prove our main theorem, we require a nonstandard version of the hyperplane separation theorem, which we give here. For $a, b \in \mathbb{R}^k$ for some finite k, let $\langle a, b\rangle$ denote the inner product. We begin by stating the standard hyperplane separation theorem:

Theorem 7.2.18 (Hyperplane separation theorem)**.** *For any $k \in \mathbb{N}$, let S_1 and S_2 be two disjoint convex subsets of $\mathbb{R}^k$, then there exists $w \in \mathbb{R}^k \setminus \{\mathbf{0}\}$ such that, for all $p_1 \in S_1$ and $p_2 \in S_2$, we have $\langle w, p_1\rangle \geq \langle w, p_2\rangle$.*

Using a suitable encoding of this theorem in first-order logic, the transfer principle yields a hyperfinite version:

Theorem 7.2.19. *Fix any $K \in {}^*\mathbb{N}$. If S_1, S_2 are two disjoint internal convex subsets of $I({}^*\mathbb{R}^K)$, then there exists $W \in \mathrm{I}({}^*\mathbb{R}^K) \setminus \{\mathbf{0}\}$ such that, for all $P_1 \in S_1$ and $P_2 \in S_2$, we have $\langle W, P_1\rangle \geq \langle W, P_2\rangle$.*

Proof. We first restate the standard hyperplane separation theorem. We shall view the set $\mathbb{R}^{\mathbb{N}}$ as the set of functions from $\mathbb{N}$ to $\mathbb{R}$. For every element $x \in \mathbb{R}^{\mathbb{N}}$, we use $x(k)$ to denote the value of the k-th coordinate of x for any $k \in \mathbb{N}$. The standard hyperplane separation theorem is equivalent to:

> For any two disjoint convex $S_1, S_2 \in \mathscr{P}(\mathbb{R}^{\mathbb{N}})$, if $\exists k \in \mathbb{N}$ such that $\forall s \in S_1 \cup S_2\ \forall k' > k$ we have $s(k') = 0$ then $\exists a \in \mathbb{R}^{\mathbb{N}} \setminus \{\mathbf{0}\}$ with $a(k') = 0$ for all $k' > k$ such that $\forall p_1 \in S_1, p_2 \in S_2$ $((\forall k' > k, a(k') = 0) \wedge (\langle a, p_1\rangle \leq \langle a, p_2\rangle))$.

By the transfer principle, we know that ${}^*(\mathbb{R}^{\mathbb{N}})$ denotes the set of all internal functions from ${}^*\mathbb{N}$ to ${}^*\mathbb{R}$. We shall view the inner product $\langle\cdot,\cdot\rangle$ to be a function from $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ to $\mathbb{R}$. Note that $\forall p, s \in \mathbb{R}^{\mathbb{N}}$ if $\exists k \in \mathbb{N}$ such that $\forall k' > k$ we have $s(k') = 0$ then $\langle p, s\rangle = \sum_{i=1}^{k} p(i)s(i)$. Thus the nonstandard extension of $\langle\cdot,\cdot\rangle$ is a function from ${}^*(\mathbb{R}^{\mathbb{N}}) \times {}^*(\mathbb{R}^{\mathbb{N}})$ to ${}^*\mathbb{R}$ satisfying the same property.

Now by the transfer principle we know that:

For any two disjoint convex sets $S_1, S_2 \in {}^*\mathscr{P}(\mathbb{R}^{\mathbb{N}})$. If $\exists K \in {}^*\mathbb{N}$ such that $\forall s \in S_1 \cup S_2$ $\forall K' > K$ we have $s(K') = 0$ then $\exists W \in {}^*(\mathbb{R}^{\mathbb{N}}) \setminus \{\mathbf{0}\}$ such that for all $p_1 \in S_1, p_2 \in S_2$ we have $((\forall K' > K, W(K') = 0) \wedge \sum_{i=1}^{K} W(i)p_1(i) \leq \sum_{i=1}^{K} W(i)p_2(i))$.

In this sentence, it is easy to see that we can view the projections of S_1, S_2 as internal subsets of $\mathrm{I}({}^*\mathbb{R}^K)$ and the projection of W as an element from $\mathrm{I}({}^*\mathbb{R}^K) \setminus \{\mathbf{0}\}$. Hence we have that: $\forall K \in {}^*\mathbb{N}$, if S_1, S_2 are two disjoint internal convex subsets of $\mathrm{I}({}^*\mathbb{R}^K)$, then there exists $W \in \mathrm{I}({}^*\mathbb{R}^K) \setminus \{\mathbf{0}\}$ such that for any $P_1 \in S_1$ and any $P_2 \in S_2$, $\sum_{i=1}^{K} W(i)P_1(i) \leq \sum_{i=1}^{K} W(i)P_2(i)$. Thus we have the desired result. □

Recall that our nonstandard model is κ-saturated for some infinite κ.

Theorem 7.2.20. *Let $\mathscr{C} \subseteq {}^{\sigma}\mathscr{D}$ be a (necessarily finite or external) set with cardinality less than κ, and suppose that $\mathscr{C}$ is a $^\circ$essentially complete subclass of $(\mathscr{C})_{FC}$. Let $\Delta_0 \in {}^*\mathscr{D}$ and suppose Δ_0 is *extended admissible on Θ among $\mathscr{C}$. Then, for every hyperfinite set $T \subseteq {}^*\Theta$ containing Θ, Δ_0 is nonstandard Bayes among $(\mathscr{C})_{FC}$ with respect to some nonstandard prior concentrating on T.*

Proof. Without loss of generality we may take $T = T_\Theta$. By Lemma 7.1.10 and the fact that $\mathscr{C}$ is an $^\circ$essentially complete subclass of $(\mathscr{C})_{FC}$, Δ_0 is *extended admissible on Θ among $(\mathscr{C})_{FC}$. By Lemma 7.1.5, Δ_0 is $\frac{1}{n}$-*admissible on T_Θ among $(\mathscr{C})_{FC}$ for every $n \in \mathbb{N}$. Hence, by Lemma 7.2.16, $Q(\Delta_0)_n \cap \mathscr{S}^{(\mathscr{C})_{FC}} = \emptyset$ for all $n \in \mathbb{N}$.

By the definition of $(\mathscr{C})_{FC}$, we have $Q(\Delta_0)_n \cap \mathscr{S}^{*\mathrm{conv}(D)} = \emptyset$ for every $D \in \mathscr{C}^{[<\infty]}$. By Lemmas 7.2.11 and 7.2.15, $\mathscr{S}^{*\mathrm{conv}(D)}$ and $Q(\Delta_0)_n$ are both internal convex sets, hence, by Theorem 7.2.19, there is a nontrivial hyperplane $\Pi_n^D \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})$ that separates them.

For every $D \in \mathscr{C}^{[<\infty]}$ and $n \in \mathbb{N}$, let $\phi_n^D(\Pi)$ be the formula

$$(\Pi \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})) \wedge (\Pi \neq \mathbf{0} \wedge (\forall x \in Q(\Delta_0)_n)\,(\forall s \in \mathscr{S}^{*\mathrm{conv}(D)})\,\langle \Pi, x \rangle \leq \langle \Pi, s \rangle), \tag{7.2.19}$$

and let $\mathscr{F} = \{\phi_n^D(\Pi) : n \in \mathbb{N}, D \in \mathscr{C}^{[<\infty]}\}$. By the above argument and the fact that $\mathscr{C}^{[<\infty]}$ is closed under taking finite unions and the sets $Q(\Delta_0)_n$, for $n \in \mathbb{N}$, are nested, $\mathscr{F}$ is finitely satisfiable. Note that $\mathscr{F}$ has cardinality no more than κ, yet our nonstandard extension is κ-saturated by hypothesis. Therefore, by the saturation principle, there exists a nontrivial hyperplane Π satisfying every sentence in $\mathscr{F}$ simultaneously. That is, there exists $\Pi \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta})$ such that $\Pi \neq \mathbf{0}$ and, for all $x \in \bigcup_{n \in \mathbb{N}} Q(\Delta_0)_n$ and for all $s \in \bigcup_{D \in \mathscr{C}^{[<\infty]}} \mathscr{S}^{*\mathrm{conv}(D)} = \mathscr{S}^{(\mathscr{C})_{FC}}$, we have $\langle \Pi, x \rangle \leq \langle \Pi, s \rangle$.

Hence, by Lemma 7.2.17, the normalized vector $\Pi / \|\Pi\|_1$ is well-defined and induces a probability measure π on ${}^*\Theta$ concentrating on T_Θ, and Δ_0 is nonstandard Bayes under π among $(\mathscr{C})_{FC}$. □

Theorem 7.2.21. *For $\delta_0 \in \mathscr{D}$, the following are equivalent statements:*

1. *δ_0 is extended admissible among $\mathscr{D}_{0,FC}$.*

2. *${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{D}_{0,FC}$.*

3. *${}^*\delta_0$ is nonstandard Bayes among $({}^\sigma\mathscr{D}_0)_{FC}$.*

If (LC) also holds, then the following statements are also equivalent:

4. *δ_0 is extended admissible among $\mathscr{D}_0$.*

5. *${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{D}_0$.*

Moreover, statements (2), (3), and (5) can be taken to assert that, for all hyperfinite sets $T \subseteq {}^*\Theta$ containing Θ, Bayes optimality holds with respect to some nonstandard prior concentrating on T.

Proof. From (1) and Theorem 7.1.8, ${}^*\delta_0$ is *extended admissible on Θ among ${}^\sigma\mathscr{D}_{0,FC}$. It follows from Lemma 7.2.14 and Theorem 7.2.20 that, for all hyperfinite sets $T \subseteq {}^*\Theta$ containing Θ, ${}^*\delta_0$ is nonstandard Bayes among $({}^\sigma\mathscr{D}_0)_{FC}$ with respect to some nonstandard prior π concentrating on T. Hence (3) holds and (2) follows trivially.

From (2) and Theorem 7.2.7, it follows that ${}^*\delta_0$ is *extended admissible on ${}^*\Theta$ among ${}^\sigma\mathscr{D}_{0,FC}$. Then (1) follows from Theorem 7.1.8.

It is the case that (1) implies (4) by Lemma 7.1.5, and the other direction follows from (LC), Lemma 6.1.13, and Lemma 6.1.4. Similarly, (2) implies (5). Finally, from (5) and Theorem 7.2.7, it follows that ${}^*\delta_0$ is *extended admissible on ${}^*\Theta$ among ${}^\sigma\mathscr{D}_0$. Then (4) follows from Theorem 7.1.8. □

It follows immediately that the class of extended admissible procedures is a complete class if and only if the class of procedures whose extensions are nonstandard Bayes are a complete class.

Remark 7.2.22. ${}^\sigma\mathscr{D}_0$, ${}^\sigma\mathscr{D}_{0,FC}$, and $({}^\sigma\mathscr{D}_0)_{FC}$ are all external. However, our model is more saturated than the external cardinalities of ${}^\sigma\mathscr{D}_0$ and ${}^\sigma\mathscr{D}_{0,FC}$, as these sets are standard-part copies of standard sets. Therefore, Lemma 7.2.2 implies an equivalence also when ${}^*\delta_0$ is ε-*Bayes among ${}^\sigma\mathscr{D}_{0,FC}$ for some $\varepsilon \approx 0$, and when ${}^*\delta_0$ is ε-*Bayes among ${}^\sigma\mathscr{D}_0$ for some $\varepsilon \approx 0$, under (LC).

Chapter 8

Push-down Results and Examples

Having established the equivalence between extended admissibility and nonstandard Bayes optimality in the previous chapter, in this chapter, we look at several implications of this result which suggests that nonstandard analysis may yield other connections between Bayesian and frequentist optimality.

In Section 8.1, we apply the nonstandard theory to obtain a standard result: assuming the parameter space is compact and risk functions are continuous, the nonstandard extension of a decision procedure is nonstandard Bayes if and only if the decision procedure itself is Bayes. Hence, when the parameter space is compact and risk functions are continuous, a decision procedure is extended admissible if and only if it is Bayes.

In Section 8.2, we employ the results of the previous section to connect admissibility and nonstandard Bayes optimality under various regularity conditions on the space and the nonstandard prior. In the process, we give a nonstandard variant of Blyth's method which gives sufficient conditions for admissibility.

In Section 8.3, we study several simple statistical decision problems to highlight the nonstandard theory and its connections to the standard theory. In Example 8.3.4, we demonstrate the equivalence between extended admissible and nonstandard Bayes in a nonparametric problem. In Example 8.3.5, we give an example of a nonstandard Bayes but not standard Bayes decision procedure. Finally, we close with some remarks and open problems in Section 8.4.

8.1 Applications to Statistical Decision Problems with Compact Parameter Space

In this section, we use our nonstandard theory to prove that, under the additional hypotheses that Θ is compact (and thus normal) and all risk functions are continuous, the class of extended admissible procedures is precisely the class of Bayes procedures. The strength of our result lies in the absence of any additional assumptions on the loss or model.[1]

Assume ${}^*\delta$ is nonstandard Bayes with respect to some nonstandard prior π on ${}^*\Theta$. In this section, we will construct a standard probability measure π_p on Θ from π in such a way that the internal risk of ${}^*\delta$ under π is infinitesimally close to the risk of δ under π_p. This then implies that π is Bayes with respect to π_p, and yields a standard characterization of extended admissible procedures.

Extension allows us to associate an internal probability measure ${}^*\pi$ to every standard probability measure π. The next theorem describes a reverse process via Loeb measures.

Lemma 8.1.1 ([11, Thm. 13.4.1])**.** *Let Y be a compact Hausdorff space equipped with Borel σ-algebra $\mathscr{B}[Y]$, let ν be an internal probability measure defined on $({}^*Y, {}^*\mathscr{B}[Y])$, and let $\mathscr{C} = \{C \subset Y : \mathsf{st}^{-1}(C) \in \overline{{}^*\mathscr{B}[Y]}_\nu\}$. Define a probability measure ν_p on the sets $\mathscr{C}$ by $\nu_p(C) = \overline{\nu}(\mathsf{st}^{-1}(C))$. Then $(Y, \mathscr{C}, \nu_p)$ is the completion of a regular Borel probability space.*

Note that $\mathsf{st}^{-1}(E)$ is Loeb measurable for all $E \in \mathscr{B}[Y]$ by Theorem 2.3.9.

Definition 8.1.2. The probability measure $\nu_p : \mathscr{C} \to [0,1]$ in Lemma 8.1.1 is called the *pushdown* of the internal probability measure ν.

Example 8.1.3. If a nonstandard prior concentrates on finitely many points in $\mathrm{NS}({}^*\Theta)$, then its pushdown concentrates on the standard parts of those points, hence is a standard measure with support on a finite set.

Example 8.1.4. Suppose $S = [K^{-1}, 2K^{-1}, \dots, 1-K^{-1}, 1]$ for some nonstandard natural $K \in {}^*\mathbb{N} \setminus \mathbb{N}$. Define an internal probability measure π on ${}^*[0,1]$ by $\pi\{s\} = K^{-1}$ for all $s \in S$, and let π_p be its pushdown. Then π_p is Lebesgue measure on $[0,1]$.

The following lemma establishes a close link between Loeb integration and integration with respect to the pushdown measure.

[1] In Section 7.2, the Hausdorff condition can be sidestepped by adopting the discrete topology. Unless Θ is finite, however, Θ will not be compact under the discrete topology. Thus, the topological hypotheses in this section not only determine the space of priors, but also restrict the set of decision problems to which the theory applies.

Lemma 8.1.5. *Let Y be a compact Hausdorff space equipped with Borel σ-algebra $\mathscr{B}[Y]$, let ν be an internal probability measure on $({}^*Y, {}^*\mathscr{B}[Y])$, let ν_p be the pushdown of ν, and let $f : Y \to \mathbb{R}$ be a bounded measurable function. Define $g : {}^*Y \to \mathbb{R}$ by $g(s) = f({}^\circ s)$. Then we have $\int f \mathrm{d}\nu_p = \int g \mathrm{d}\overline{\nu}$.*

Proof. 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 $\mathscr{F}_n = \{F_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ forms a finite partition of Y, and similarly for $\mathscr{G}_n = \{G_{n,k} : k \in \mathbb{Z}\} \setminus \{\emptyset\}$ and ${}^*Y$. For every $n \in \mathbb{N}$, define $\hat{f}_n : Y \to \mathbb{R}$ and $\hat{g}_n : {}^*Y \to \mathbb{R}$ 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 $\mathscr{F}_n$ (resp., $\mathscr{G}_n$). By construction $\hat{f}_n \le f < \hat{f}_n + \frac{1}{n}$ and $\hat{g}_n \le g < \hat{g}_n + \frac{1}{n}$. Note that $G_{n,k} = \mathsf{st}^{-1}(F_{n,k})$ for every $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. Moreover, Y is even regular Hausdorff, hence Lemma 2.4.10 implies that $G_{n,k}$ is $\overline{\nu}$-measurable. It follows that $\int f \mathrm{d}\nu_p = \lim_{n\to\infty} \int \hat{f}_n \mathrm{d}\nu_p$ and $\int g \mathrm{d}\overline{\nu} = \lim_{n\to\infty} \int \hat{g}_n \mathrm{d}\overline{\nu}$. Moreover, by Lemma 8.1.1, 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 $\int \hat{f}_n \mathrm{d}\nu_p = \int \hat{g}_n \mathrm{d}\overline{\nu}$. Hence we have $\int g \mathrm{d}\overline{\nu} = \int f \mathrm{d}\nu_p$, completing the proof. □

In order to control the difference between the internal and standard Bayes risks under a nonstandard prior π and its pushdown π_p, it will suffice to require that risk functions be continuous. (Recall that we quoted results listing natural conditions that imply continuous risk in Theorems 6.2.4 and 6.2.5.)

Condition RC (risk continuity)**.** $r(\cdot, \delta)$ is continuous on Θ, for all $\delta \in \mathscr{D}$.

In order to understand the nonstandard implications of this regularity condition, we introduce the following definition from nonstandard analysis.

Definition 8.1.6. Let X and Y be topological spaces. A function $f : {}^*X \to {}^*Y$ is S-continuous at $x \in {}^*X$ if $f(y) \approx f(x)$ for all $y \approx x$.

A fundamental result in nonstandard analysis links continuity and S-continuity:

Lemma 8.1.7. *Let X and Y be Hausdorff spaces, where Y is also locally compact, and let $D \subseteq X$. If a function $f : X \to Y$ is continuous on D then its extension ${}^*f$ is* $\mathrm{NS}({}^*Y)$*-valued and S-continuous on* $\mathrm{NS}({}^*D)$.

See **??** for a proof of this classical result. We are now at the place to establish the correspondence between internal Bayes risk and standard Bayes risk. The proof relies on the following technical lemma.

Lemma 8.1.8 ([3, Cor. 4.6.1])**.** *Suppose $(\Omega, \mathscr{F}, P)$ is an internal probability space, and $F : \Omega \to {}^*\mathbb{R}$ is an internal P-integrable function such that ${}^\circ F$ exists everywhere. Then ${}^\circ F$ is integrable with respect to $\overline{P}$ and $\int F \mathrm{d}P \approx \int {}^\circ F \mathrm{d}\overline{P}$.*

Lemma 8.1.9. *Suppose Θ is compact Hausdorff and (RC) holds. Let π be an internal distribution on ${}^*\Theta$ and let $\pi_p : \mathscr{C} \to [0,1]$ be its pushdown. Let $\delta_0 \in \mathscr{D}$ be a standard decision procedure. If ${}^*r(\cdot, {}^*\delta_0)$ is π-integrable then $r(\cdot, \delta_0)$ is a π_p-integrable function and $r(\pi_p, \delta_0) \approx {}^*r(\pi, {}^*\delta_0)$, i.e., the Bayes risk under π_p of δ_0 is within an infinitesimal of the nonstandard Bayes risk under π of ${}^*\delta_0$.*

Proof. Because Θ is compact Hausdorff, ${}^\circ t$ exists for all $t \in {}^*\Theta$ and Lemma 8.1.1 implies π_p is a probability measure on $(\Theta, \mathscr{C})$, where $\mathscr{C}$ is the π_p-completion of $\mathscr{B}[\Theta]$. By (RC) and Lemma 8.1.7, for all $t \in {}^*\Theta$, we have

$$ {}^*r(t, {}^*\delta_0) \approx {}^*r({}^\circ t, {}^*\delta_0) = r({}^\circ t, \delta_0). \tag{8.1.1} $$

Hence ${}^\circ({}^*r(t, \delta_0)) = r({}^\circ t, \delta_0)$ exists for all $t \in {}^*\Theta$. As ${}^*r(\cdot, {}^*\delta_0)$ is π-integrable, by Lemma 8.1.8, we know that ${}^\circ({}^*r(\cdot, {}^*\delta_0))$ is $\overline{\pi}$-integrable and

$$ \int {}^*r(t, {}^*\delta_0)\pi(\mathrm{d}t) \approx \int {}^\circ({}^*r(t, {}^*\delta_0))\overline{\pi}(\mathrm{d}t) = \int {}^*r({}^\circ t, {}^*\delta_0)\overline{\pi}(\mathrm{d}t). \tag{8.1.2} $$

By (RC) and the fact that Θ is compact, it follows that $r(\cdot, \delta_0)$ is bounded. Thus, by Lemma 8.1.5, $\int {}^*r({}^\circ t, {}^*\delta_0)\overline{\pi}(\mathrm{d}t) = \int r(\theta, \delta_0)\pi_p(\mathrm{d}\theta)$, completing the proof. ☐

Lemma 8.1.10. *Suppose Θ is compact Hausdorff and (RC) holds. Let $\delta_0 \in \mathscr{D}$ and $\mathscr{C} \subseteq \mathscr{D}$. If ${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$, then δ_0 is Bayes among $\mathscr{C}$.*

Proof. By Theorem 7.2.21, we may assume that ${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$ with respect to a nonstandard prior π that concentrates on some hyperfinite set T. Let $\delta \in \mathscr{C}$. Then ${}^*\delta \in {}^\sigma\mathscr{C}$, hence ${}^*r(\pi, {}^*\delta_0) \lessapprox {}^*r(\pi, {}^*\delta)$. Let π_p denote the pushdown of π. As Θ is compact Hausdorff, we know that π_p is a probability measure. As π concentrates on the hyperfinite set T, we know that ${}^*r(\cdot, {}^*\delta_0)$ and ${}^*r(\cdot, {}^*\delta)$ are π-integrable. By Lemma 8.1.9, we have $r(\pi_p, \delta_0) \approx {}^*r(\pi, {}^*\delta_0)$ and $r(\pi_p, \delta) \approx {}^*r(\pi, {}^*\delta)$. Thus, we know that $r(\pi_p, \delta_0) \leq r(\pi_p, \delta)$. As our choice of δ was arbitrary, δ_0 is Bayes under π_p among $\mathscr{C}$. ☐

Theorem 8.1.11. *Suppose Θ is compact Hausdorff and (RC) holds. For $\delta_0 \in \mathscr{D}$, the following statements are equivalent:*

1. *δ_0 is extended admissible among $\mathscr{D}_{0,FC}$.*
2. *δ_0 is extended Bayes among $\mathscr{D}_{0,FC}$.*
3. *δ_0 is Bayes among $\mathscr{D}_{0,FC}$.*

If (LC) also holds, then the equivalence extends to these statements with $\mathscr{D}_0$ in place of $\mathscr{D}_{0,FC}$.

Proof. Suppose (1) holds. Then by Theorem 7.2.21, ${}^*\delta_0$ is nonstandard Bayes among ${}^\sigma\mathscr{D}_{0,FC}$. Then (3) follows from Lemma 8.1.10. The reverse implications follows from Theorem 6.1.8.

The statements with $\mathscr{D}_{0,FC}$ imply those for $\mathscr{D}_0 \subseteq \mathscr{D}_{0,FC}$ trivially. When (LC) holds, we have Lemma 6.1.13. Hence, the reverse implications follows from Lemma 6.1.4 and Theorem 6.1.9. □

We conclude this section with a strengthening of Theorem 7.2.21, showing that infinitesimal *Bayes risk yields zero *Bayes risk, and that a procedure is optimal among all extensions if and only if it optimal among all internal estimators:

Corollary 8.1.12. *Suppose Θ is compact Hausdorff and (RC) holds. For $\delta_0 \in \mathscr{D}$, the following statements are equivalent:*

1. *δ_0 is extended admissible among $\mathscr{D}_{0,FC}$.*

2. *${}^*\delta_0$ is nonstandard Bayes among ${}^*\mathscr{D}_{0,FC}$.*

3. *${}^*\delta_0$ is 0-*Bayes among ${}^*\mathscr{D}_{0,FC}$.*

Moreover, the equivalence extends to these statements with ${}^\sigma\mathscr{D}_{0,FC}$ in place of ${}^*\mathscr{D}_{0,FC}$. If (LC) also holds, the equivalence extends to these statement with $\mathscr{D}_0/{}^\sigma\mathscr{D}_0/{}^*\mathscr{D}_0$ in place of $\mathscr{D}_{0,FC}/{}^\sigma\mathscr{D}_{0,FC}/{}^*\mathscr{D}_{0,FC}$.

Proof. Statement (1) implies that δ_0 is Bayes among $\mathscr{D}_{0,FC}$ by Theorem 8.1.11. This implies (3) by transfer, (3) implies (2) by definition, and (2) implies (1) by Theorem 7.2.21.

Statements (2) and (3) with ${}^*\mathscr{D}_{0,FC}$ imply their counterparts with ${}^\sigma\mathscr{D}_{0,FC}$ in place of ${}^*\mathscr{D}_{0,FC}$, trivially. Statement (3) with ${}^\sigma\mathscr{D}_{0,FC}$ implies (2) with ${}^\sigma\mathscr{D}_{0,FC}$ which implies (1) by Theorem 7.2.21.

The additional equivalences under (LC) follow by the same logic as above and in the proof of Theorem 7.2.21. □

8.2 Admissibility of Nonstandard Bayes Procedures

Heretofore, we have focused on the connection between extended admissibility and nonstandard Bayes optimality. In this section, we shift our focus to the admissibility of decision procedures whose extensions are nonstandard Bayes. In all but the final result of this section, *we will assume that Θ is a metric space* and write d for the metric.

On finite parameter spaces with bounded loss, it is known that Bayes procedures with respect to priors assigning positive mass to every state are admissible. Similarly, when risk functions are continuous, Bayes procedures with respect to priors with full support are admissible. We can establish analogues of these result on general parameter spaces by a suitable nonstandard relaxation of a standard prior having full support.

Definition 8.2.1. For $x, y \in {}^*\mathbb{R}$, write $x \gg y$ when $\gamma x > y$ for all $\gamma \in \mathbb{R}_{>0}$.

Definition 8.2.2. Let X be a metric space with metric d, and let $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$. An internal probability measure π on ${}^*\Theta$ is *ε-regular* if, for every $\theta_0 \in \Theta$ and non-infinitesimal $r > 0$, we have $\pi(\{t \in {}^*\Theta : {}^*d(t, \theta_0) < r\}) \gg \varepsilon$.

The following result establishes *admissibility from *Bayes optimality under conditions analogues to full support and continuity of the risk function.

Lemma 8.2.3. *Suppose Θ is a metric space. Let $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$, $\Delta_0 \in {}^*\mathscr{D}$, and $\mathscr{C} \subseteq {}^*\mathscr{D}$, and suppose ${}^*r(\cdot, \Delta)$ is S-continuous on* $\mathrm{NS}({}^*\Theta)$ *for all $\Delta \in \mathscr{C} \cup \{\Delta_0\}$.*

If Δ_0 is ε-*Bayes among $\mathscr{C}$ with respect to an ε-regular nonstandard prior π, then Δ_0 is *admissible in $\Theta/{}^*\Theta$ among $\mathscr{C}$.

Proof. Suppose Δ_0 is not *admissible in $\Theta/{}^*\Theta$ among $\mathscr{C}$. Then, for some $\Delta \in \mathscr{C}$ and $\theta_0 \in \Theta$, it holds that

$$(\forall \theta \in {}^*\Theta)({}^*r(\theta, \Delta) \leq {}^*r(\theta, \Delta_0)) \tag{8.2.1}$$

$$\text{and} \quad {}^*r(\theta_0, \Delta) \not\approx {}^*r(\theta_0, \Delta_0). \tag{8.2.2}$$

From Eq. (8.2.2), ${}^*r(\theta_0, \Delta_0) - {}^*r(\theta_0, \Delta) > 2\gamma$ for some positive $\gamma \in \mathbb{R}$. Let A be the set of all $a \in {}^*\mathbb{R}_{>0}$ such that

$$(\forall t \in {}^*\Theta)\ ({}^*d(t, \theta_0) < a \implies {}^*r(t, \Delta_0) - {}^*r(t, \Delta) > \gamma). \tag{8.2.3}$$

By the S-continuity of ${}^*r$ on $\mathrm{NS}({}^*\Theta)$, the set A contains all infinitesimals. By saturation and the fact that A is an internal set, A must contain some positive $a_0 \in \mathbb{R}$. In summary,

$$(\forall t \in {}^*\Theta)\ ({}^*d(t, \theta_0) < a_0 \implies {}^*r(t, \Delta_0) - {}^*r(t, \Delta) > \gamma). \tag{8.2.4}$$

Let $M = \{t \in {}^*\Theta : {}^*d(t, \theta_0) < a_0\}$. By the internal definition principle, M is an internal set. By Eq. (8.2.1) and the definition and internality of M, the difference in internal Bayes risk between Δ_0 and Δ satisfies

$$
\begin{aligned}
{}^*r(\pi, \Delta_0) - {}^*r(\pi, \Delta) &= \int_{{}^*\Theta} ({}^*r(t, \Delta_0) - {}^*r(t, \Delta))\pi(\mathrm{d}t) &\quad (8.2.5)\\
&\geq \int_M ({}^*r(t, \Delta_0) - {}^*r(t, \Delta))\pi(\mathrm{d}t) > \gamma\pi(M). &\quad (8.2.6)
\end{aligned}
$$

But $\gamma\pi(M) > \varepsilon$ because π is ε-regular, hence Δ_0 is not ε-*Bayes among $\mathscr{C}$ with respect to π. □

The following theorem is an immediate consequence of Lemma 8.2.3 and is a nonstandard analogue of Blyth's Method [26, §5 Thm. 7.13] (see also [26, §5 Thm. 8.7]). In Blyth's method, a sequence of (potentially improper) priors with sufficient support is used to establish the admissibility of a decision procedure. In contrast, a single nonstandard prior witnesses the nonstandard admissibility of a nonstandard Bayes procedure.

Theorem 8.2.4. *Suppose Θ is a metric space and (RC) holds. Let $\delta_0 \in \mathscr{D}$ and $\mathscr{C} \subset \mathscr{D}$. If there exists $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$ such that ${}^*\delta_0$ is ε-*Bayes among ${}^\sigma\mathscr{C}$ with respect to an ε-regular nonstandard prior π, then ${}^*\delta_0$ is *admissible in $\Theta/{}^*\Theta$ among ${}^\sigma\mathscr{C}$.*

Proof. By (RC) and Lemma 8.1.7, for all $\delta \in \mathscr{D}$, $\theta_0 \in \Theta$, and $t \approx \theta_0$, we have ${}^*r(t, {}^*\delta) \approx {}^*r(\theta_0, {}^*\delta)$. By Lemma 8.2.3, ${}^*\delta_0$ is *admissible in $\Theta/{}^*\Theta$ among ${}^\sigma\mathscr{C}$. □

These theorems have the following consequence for standard decision procedures:

Theorem 8.2.5. *Suppose Θ is a metric space and (RC) holds, and let $\delta_0 \in \mathscr{D}$ and $\mathscr{C} \subseteq \mathscr{D}$. If there exists $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$ such that ${}^*\delta_0$ is ε-*Bayes among ${}^\sigma\mathscr{C}$ with respect to an ε-regular nonstandard prior, then δ_0 is admissible among $\mathscr{C}$.*

Proof. The result follows from Theorem 7.1.7 and Theorem 8.2.4. □

Theorem 8.2.5 implies the well-known result that Bayes procedures with respect to priors with full support are admissible [14, §2.3 Thm. 3] (see also [26, §5 Thm. 7.9]).

Theorem 8.2.6. *Suppose Θ is a metric space and (RC) holds and let $\delta_0 \in \mathscr{D}$. If δ_0 is Bayes among $\mathscr{D}$ with respect to a prior π with full support, then δ_0 is admissible among $\mathscr{D}$.*

Proof. Note that δ_0 is Bayes under π among $\mathscr{D}$ if and only if ${}^*\delta_0$ is nonstandard Bayes under ${}^*\pi$ among ${}^\sigma\mathscr{D}$. As π has full support, ${}^*\pi$ is ε-regular for every infinitesimal $\varepsilon \in {}^*\mathbb{R}_{>0}$. By Theorem 8.2.5, we have the desired result. ☐

We close with an admissibility result requiring no additional regularity:

Theorem 8.2.7. *Let $\delta_0 \in \mathscr{D}$ and $\mathscr{C} \subseteq \mathscr{D}$. If there exists $\varepsilon \in {}^*\mathbb{R}_{\geq 0}$ such that ${}^*\delta_0$ is ε-*Bayes among ${}^*\mathscr{C}$ with respect to a nonstandard prior π satisfying $\pi\{\theta\} \gg \varepsilon$ for all $\theta \in \Theta$, then δ_0 is admissible among $\mathscr{C}$.*

Proof. Suppose δ_0 is not admissible among $\mathscr{C}$. Then by Theorem 7.1.7, ${}^*\delta_0$ is not *admissible in $\Theta/{}^*\Theta$ among ${}^\sigma\mathscr{C}$. Thus there exists $\delta \in \mathscr{C}$ and $\theta_0 \in \Theta$ such that ${}^*r(\theta, {}^*\delta) \leq {}^*r(\theta, {}^*\delta_0)$ for all $\theta \in {}^*\Theta$ and ${}^*r(\theta_0, {}^*\delta_0) - {}^*r(\theta_0, {}^*\delta) > \gamma$ for some $\gamma \in \mathbb{R}_{>0}$. Then ${}^*r(\pi, {}^*\delta_0) - {}^*r(\pi, {}^*\delta) \geq \pi\{\theta_0\}\gamma > \varepsilon$. But this implies that ${}^*\delta_0$ is not ε-*Bayes under π among $\mathscr{C}$. ☐

Remark 8.2.8. The astute reader may notice that Theorem 8.2.7 is actually a corollary of Theorem 8.2.5 provided we adopt the discrete topology/metric on Θ. Changing the metric changes the set of available prior distributions and also changes the set of ε-regular nonstandard priors. See also Remark 8.3.3.

8.3 Some Examples

The following examples serve to highlight some of the interesting properties of our nonstandard theory and its consequences for classical problems.

Example 8.3.1. Consider any standard statistical decision problem with a finite, discrete (hence compact) parameter space. (RC) holds trivially, and so Theorem 8.1.11 and Corollary 8.1.12 imply that a decision procedure is extended admissible if and only if it is extended Bayes if and only if it is Bayes if and only if its extension is nonstandard Bayes among all internal decision procedures. By Theorem 8.2.6, we obtain another classical result: if a procedure is Bayes with respect to a prior with full support, it is admissible.

Example 8.3.2. Consider the classical problem of estimating the mean of a multivariate normal distribution in d dimensions under squared error when the covariance matrix is known to be the identity matrix. By the convexity of the squared error loss function, Lemma 6.1.13 implies that the nonrandomized procedures form an essentially complete class. (Indeed, the loss is strictly convex and so the nonrandomized procedures are a complete class.) Theorem 7.2.21 implies that every extended admissible estimator among $\mathscr{D}_0$ is nonstandard Bayes among ${}^{\sigma}\mathscr{D}_{0,FC}$.

We can derive further results if we can establish that risk functions are continuous. Indeed, one can use Theorem 6.2.4 to establish that (RC) holds in the normal-location problem. Theorem 8.2.6 then implies that every Bayes estimator with respect to a prior with full support is admissible. In particular, for every $k > 0$, the estimator $\delta_k^B(\mathbf{x}) = \frac{k^2}{k^2+1}\mathbf{x}$ is Bayes with respect to the full-support prior $\pi_k = \mathscr{N}(0, k^2 I_d)$, hence admissible.

Consider now the maximum likelihood estimator $\delta^M(\mathbf{x}) = \mathbf{x}$ and let K be an infinite natural number. Then ${}^*\delta^M(\mathbf{x}) \approx ({}^*\delta^B)_K(\mathbf{x})$ for all $\mathbf{x} \in \mathrm{NS}({}^*\mathbb{R}^d)$, where ${}^*\delta^B$ is the extension of the function $k \mapsto \delta_k^B$. The normal prior $({}^*\pi)_K$ is "flat" on $\mathbb{R}^d$ in the sense that, at every near-standard real number, the ratio of the density to $(2\pi)^{-\frac{d}{2}}K^{-d}$ is within an infinitesimal of 1. These observations provide a nonstandard interpretation to the idea that the maximum likelihood estimator is a Bayes estimator given a "uniform" prior.

Since (RC) holds, Theorem 8.2.5 implies that every estimator whose extension is ε-*Bayes among ${}^{\sigma}\mathscr{D}_0$ with respect to an ε-regular prior is admissible among $\mathscr{D}_0$. An easy calculation reveals that the Bayes risk of $({}^*\delta^B)_K$ with respect to $({}^*\pi)_K$ is $d\frac{K^2}{K^2+1}$, while the Bayes risk of ${}^*\delta^M$ with respect to $({}^*\pi)_K$ is d. Thus, ${}^*\delta^M$ is even nonstandard Bayes among ${}^*\mathscr{D}$, and in particular, ${}^*\delta^M$ is ε-*Bayes under $({}^*\pi)_K$ among ${}^*\mathscr{D}$ for $\varepsilon = d(K^2+1)^{-1}$. From the density above, it is then straightforward to verify that, for $d = 1$, the prior $({}^*\pi)_K$ is ε-regular, but that it fails to be for $d \geq 2$. Therefore, by Theorem 8.2.5, it follows that δ^M is admissible among $\mathscr{D}_0$ for $d = 1$, as is well known. The theorem is silent in this case for $d \geq 2$. Indeed, Stein [50] famously showed that δ^M is admissible for $d = 2$ and inadmissible for $d \geq 3$.

Remark 8.3.3. Here we have used Theorem 8.2.5 and the standard metric on $\Theta = \mathbb{R}^d$ in order to establish admissibility. Note that the infinite-variance Gaussian prior is not ε-regular with respect to the discrete metric on Θ, and so a different nonstandard prior would have been needed to establish admissibility via Theorem 8.2.7.

The next example is a simple demonstration of extended admissibility in a nonparametric estimation problem.

Example 8.3.4. Let $\Theta \subseteq \mathscr{M}_1(\mathbb{R})$ be the set of probability measures on $\mathbb{R}$ with finite first moment, and consider the model $P_\theta = \theta$ under which we observe a single sample from the unknown distribution θ of interest. Taking $\mathbb{A} = \Theta$, we would like to estimate an unknown $\theta \in \Theta$ under Wasserstein loss

$$\ell(\theta, \hat{\theta}) = \inf_{\mu} \int d(x,y)\, \mu(\mathrm{d}(x,y)) \geq 0, \tag{8.3.1}$$

where d is the standard Euclidean metric and the infimum is taken over all couplings of θ and $\hat{\theta}$, i.e., over all $\mu \in \mathscr{M}_1(\mathbb{R} \times \mathbb{R})$ with marginals θ and $\hat{\theta}$, respectively. Consider the estimator $\delta_0(x) = \mathrm{Dirac}(x)$ that degenerates on the observed sample. Let H be a *Uniform distribution on $[-k,k]$, for k infinite, and let π be a *Dirichlet process prior with *base measure αH, where $0 < \alpha \ll k^{-1}$. (We will drop the modifier * and rely on context to disambiguate whether we are referring to a standard concept or its transfer.) Let G be a random probability measure with distribution π, and, conditioned on G, let X_1, X_2 be independent random variables with distribution G. By transfer and the properties of the Dirichlet process, $\mathbb{P}\{X_1 \neq X_2\} = \frac{\alpha}{\alpha+1}$. In terms of these random variables, the average risk of $^*\delta_0$ under π is the expectation $\mathbb{E}[\ell(G, \mathrm{Dirac}(X_1))]$ and this quantity is bounded by $\frac{\alpha}{\alpha+1} k \ll 1$, hence $^*\delta_0$ is nonstandard Bayes among all *estimators, hence extended Bayes and extended admissible.

In Section 8.1, we established that class of Bayes procedures coincides with the class of extended admissible estimators under compactness of the parameter space and continuity of the risk. The next example demonstrates that extended admissibility and Bayes optimality do not necessarily align if we drop the risk continuity assumption, even when the parameter space is compact. We study a non-Bayes admissible estimator and characterize a nonstandard prior with respect to which it is nonstandard Bayes.

Example 8.3.5. Let $X = \{0,1\}$ and $\Theta = [0,1]$, the latter viewed as a subset of Euclidean space. Define $g : [0,1] \to [0,1]$ by $g(x) = x$ for $x > 0$ and $g(0) = 1$, and let $P_t = \mathrm{Bernoulli}(g(t))$, for $t \in [0,1]$, where $\mathrm{Bernoulli}(p)$ denotes the distribution on $\{0,1\}$ with mean $p \in [0,1]$. Every nonrandomized decision procedure $\delta : \{0,1\} \to [0,1]$ thus corresponds with a pair $(\delta(0), \delta(1)) \in [0,1]^2$, and so we will express nonrandomized decision procedures as pairs. Consider the loss function $\ell(x,y) = (g(x) - y)^2$. (For every x, the map $y \mapsto \ell(x,y)$ is convex but merely lower semicontinuous on $[0,1]$. It follows from Lemma 6.1.13 that nonrandomized procedures form an essentially complete class.)

Theorem 8.3.6. *In Example 8.3.5, $(0,0)$ is an admissible non-Bayes estimator.*

Proof. Let $(a,b) \in [0,1]^2$ and let $c = \min\{a,b\}$. For every $n \in \mathbb{N}$, we have

$$r(n^{-1}, (a,b)) = (1 - n^{-1})\ell(1/n, a) + n^{-1}\,\ell(1/n, b) \tag{8.3.2}$$

and so, for sufficiently large n, we have $r(n^{-1},(a,b)) \geq r(n^{-1},(c,c))$. But, for every $d > 0$ and sufficiently large n, it also holds that $r(n^{-1},(d,d)) > r(n^{-1},(0,0))$. Hence, (a,b) does not dominate $(0,0)$, hence $(0,0)$ is admissible.

To see that $(0,0)$ is not Bayes, note that an estimator (a,b) has the same Bayes risk under π as it would under the (pushforward) prior $\nu = \pi \circ g^{-1}$ in the statistical decision problem with sample space X, parameter space $\Theta' = g(\Theta) = (0,1]$, model $P'_t = \text{Bernoulli}(t)$, and squared error loss $\ell'(x,y) = (x,y)$. However, in this case, the loss is strictly convex and so the Bayes optimal decision is unique and is the posterior mean, which is a value in $(0,1]$, hence $(0,0)$ cannot be Bayes optimal for any prior. □

The failure of $(0,0)$ to be Bayes optimal is due to the fact that the posterior mean cannot be 0. However, in the nonstandard universe, the posterior mean can be made to be infinitesimal, in which case the Bayes risk of $(0,0)$ is also infinitesimal.

Theorem 8.3.7. ${}^*(0,0)$ *is nonstandard Bayes with respect to any prior concentrating on some infinitesimal* $\varepsilon > 0$.

Proof. Pick any positive infinitesimal ε and consider the nonstandard prior π concentrated on ε. The nonstandard Bayes risk of $(0,0)$ with respect to π is

$$ {}^*r(\pi,(0,0)) = {}^*r(\varepsilon,(0,0)) = \varepsilon(\varepsilon-0)^2 + (1-\varepsilon)(\varepsilon-0)^2 \approx 0 \tag{8.3.3} $$

Because the loss function in Example 8.3.5 is nonnegative, $(0,0)$ must be a nonstandard Bayes estimator with respect to π. □

We close by observing that $(0,0)$ is a generalized Bayesian estimator. In particular, the generalized Bayes risk with respect to the improper prior $\pi(\mathrm{d}\theta) = \theta^{-2}\mathrm{d}\theta$ is finite, whereas every other estimator has infinite Bayes risk. The modified statistical decision problem with parameter space $\Theta' = (0,1]$ under the standard topology, model P' and loss ℓ' meets the hypotheses of Theorem 6.2.10— indeed, the modified problem is that of estimating the mean of an exponential family model— hence every extended admissible procedure is generalized Bayes. The original problem does not meet the hypotheses of Theorem 6.2.10, since the loss is not jointly continuous.

8.4 Miscellaneous Remarks

(i) We have required Θ to be Hausdorff in order for the standard part map to be uniquely defined. Relaxing this assumption would require that we work with a standard part relation instead. At this moment, we see no roadblocks.

(ii) Assume ${}^*\delta$ is nonstandard Bayes among ${}^\sigma\mathscr{C}$. Under what conditions can we conclude that ${}^*\delta$ is ε-*Bayes among ${}^\sigma\mathscr{C}$ for all $\varepsilon \in {}^*\mathbb{R}_{>0}$? For $\varepsilon = 0$? When can we conclude δ is extended Bayes among $\mathscr{C}$? Bayes? In Corollary 8.1.12, we show that all these conditions collapse for $\mathscr{C} = \mathscr{D}_{0,FC}$ when Θ is compact Hausdorff and (RC) holds. Identifying problems that separate all of these conditions and sufficient conditions that collapse them is important work.

(iii) As a starting point towards the questions in part (ii), it is an open problem to find a procedure δ such that δ is extended admissible among $\mathscr{D}_{0,FC}$ (or even extended Bayes among $\mathscr{D}_{0,FC}$) but ${}^*\delta$ is not 0-*Bayes among ${}^\sigma\mathscr{D}_{0,FC}$. Example 8.3.2 demonstrates that δ^M is extended Bayes among $\mathscr{D}_{0,FC}$ for every dimension $d \geq 1$. A well-known variance argument shows that δ^M is also never Bayes among $\mathscr{D}_{0,FC}$, hence ${}^*\delta^M$ is never 0-*Bayes among ${}^*\mathscr{D}_{0,FC}$. The inadmissibility of δ^M in dimensions $d \geq 3$ implies that, if ${}^*\delta^M$ is 0-*Bayes among ${}^\sigma\mathscr{D}_{0,FC}$ with respect to some nonstandard prior π, then π is not a 0-regular nonstandard prior.

(iv) We restricted our attention to decision procedures whose risk functions are everywhere finite. However, if we do not make this restriction, it is possible for an admissible decision procedures to have infinite risk in some state $\theta \in \Theta$ [10, §4A.13 Part (iv)]. We make repeated use of the finite risk property and so it would be an interesting contribution to relax this assumption. A related issue is our restriction to nonnegative real-valued loss functions. It would be straightforward to allow loss functions that are bounded below or above. Allowing arbitrary loss functions, however, raises the possibility that a decision procedure's risk could be undefined on some subset of the parameter space.

(v) It is worth searching for a converse to Theorem 8.2.7, perhaps with a view to identifying a nonstandard analogue of Stein's necessary and sufficient condition for admissibility [49], but one witnessed by a single (nonstandard) prior distribution.

(vi) Our standard result, Theorem 8.1.11, is similar to Theorem 6.2.6 of Berger [4] and Theorem 6.2.3 of Wald [54]. Our theorem identifies the class of extended admissible procedures and the class of Bayes procedures, and does so by assuming that risk functions are continuous and the parameter space is compact Hausdorff. These are weaker assumptions than those of Berger, and more natural than those of Wald. It would take some work to understand which assumptions of theirs are needed to show that the extended admissible procedures (equivalently, the Bayes procedures) form a complete class. In our opinion, it is preferable to understand conditions under which we can identifying extended admissibility and Bayes optimality and then separately understand conditions under which the former is a complete class. (The classical textbook by Blackwell and Girschick [8] adopts a similar aesthetic principle.) We believe that the methods developed in this paper may allow us to remove or generalize regularity conditions in other existing results.

(vii) It would be illuminating to uncover a complete characterization of the relationships between nonstandard Bayes procedures, extended Bayes procedures, limits of Bayes procedures, and generalized Bayes procedures. Some connections can be identified simply by transfer: e.g., we already know that extended Bayes procedures are equivalent to nonstandard Bayes by a simple transfer argument and normal-form generalized Bayes implies nonstandard Bayes. Given our theorems connecting extended admissibility and nonstandard Bayes optimality, progress on this question immediately yields new connections between extended admissibility and these relaxed notions of Bayes optimality.

(viii) In general, by our main result Theorem 7.2.21, extensions of extended admissible procedures are nonstandard Bayes among the extensions of standard procedures. Under what conditions are they also nonstandard Bayes among all internal decision procedures? By Theorem 7.2.3, this is equivalent to the following natural question: When are extended admissible procedures also extended Bayes? Under the assumption of bounded risk, [8, Thm. 5.5.3] gives necessary and sufficient conditions in terms of derived two-player games. Under compactness and risk continuity, Theorem 8.1.11 shows that extended admissible procedures are even Bayes, and so of course the equivalence holds there. Identifying natural sufficient conditions or counterexamples in natural decision problems would be very enlightening.

Conclusion

In this dissertation, we studied two fundamental problems in Markov processes and statistical decision theory using nonstandard analysis. In the first case, we characterized every continuous-time Markov processes with general state space by a hyperfinite Markov process, i.e infinite Markov processes with the same first-order logic properties as finite Markov processes. By proving the ergodicity of hyperfinite Markov processes, we establish the standard Markov chain ergodic theorem for continuous-time Markov processes with general state space under moderate regularity conditions. Unlike existing results in the literature, our Markov chain ergodic theorem does not depend on drift conditions or skeleton chains. In the second case, we studied the relation between frequentist and Bayesian optimality in statistical decision theory framework, showing that a decision procedure is extended admissible if and only if it is nonstandard Bayes, i.e. it has infinitesimal excess Bayes risk. We show that this equivalence holds for arbitrary decision problems without technical conditions. Using this equivalence relation, we show that extended admissibility is equivalent to standard Bayes for decision problems with compact parameter spaces and continuous risk functions and also a nonstandard variant of Blyth method.

The idea of "hyperfinite representation" in nonstandard analysis provide a direct connection between finite mathematics and infinite mathematics. Given a infinite mathematic problem, the hyperfinite counterpart of this problem has the same first-order logic properties as the finite version of this problem. Thus, we can obtain a solution of the hyperfinite counterpart as long as we have a solution for the finite version of the problem. We then apply push-down technique to obtain the solution of the original problem. The hyperfinite Markov processes constructed from continuous-time Markov processes with general state space can be considered as hyperfinite representations of the original Markov processes since its internal transition probability differs from the transition probability of the original Markov process by only infinitesimal. Thus, many features of general Markov processes can be understood via studying its hyperfinite counterpart.

Nonstandard analysis, in many cases, offers insights to existing problems from a new prospective.

For statistical decision problem with infinite parameter spaces, one must relax the notion of Bayesian optimality to obtain connections between frequentist and Bayesian optimality. Moreover, such results are subject to regularity conditions. However, using nonstandard analysis and nonstandard probability theory, we are able to establish a link between admissibility and Bayesian optimality without technical conditions. Such surprising connection suggests that nonstandard analysis maybe able to generate fruitful result and provide new insights to mathematical statistics.

There are many avenues worth further investigation. In the proof of Theorem 8.1.11, the existence of a standard prior relies critically on compactness or, at least, on the near-standard part have Loeb-measure one. As a result, in a general non-compact setting, the pushdown measure may fail to be a probability measure, and might even be the null measure. Other notions of pushdown measure exist, although they generally produce merely finitely additive probability measures. There is a substantial body of work on finitely additive probability and its application to foundational problems in statistics and game theory [16, 21, 45–48]. It would be worthwhile identifying relationships between the nonstandard and finitely additive frameworks. Understanding such relation will enable us to answer questions such as, when does nonstandard Bayes imply finitely-additive Bayes? Other questions include but not restricted to: 1)studying particular stochastic processes (Dirichlet process, Langevin diffusion, etc) from a nonstandard prospective. 2) understanding minimaxity, stepwise Bayes, limit of Bayes by considering their hyperfinite counterpart.

As one might has already noticed, nonstandard analysis is very powerful in establishing existence result. One of many ongoing work is to show the existence of matching prior for compact parameter spaces. The idea is to first show the existence of nonstandard matching prior in the hyperfinite counterpart of this problem. As the parameter space is compact, we apply the same technique in Theorem 8.1.11 to obtain a standard marching prior.

This dissertation has explored the boundary of stochastic processes, probability theory, statistical decision theory and nonstandard analysis. In closing, we believe that nonstandard analysis will greatly enhance our understanding on connections between finite problems and infinite problems and it has much to offer to the statistical community beyond those subjects mentioned in this dissertation.

Bibliography

[1] R. M. Anderson. A nonstandard representation for Brownian motion and Itô integration. *Bull. Amer. Math. Soc.*, 82(1):99–101, 1976. ISSN 0002-9904.

[2] R. M. Anderson. Star-finite representations of measure spaces. *Trans. Amer. Math. Soc.*, 271 (2):667–687, 1982. doi: 10.2307/1998904. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.2307/1998904.

[3] L. O. Arkeryd, N. J. Cutland, and C. W. Henson, editors. *Nonstandard analysis*, volume 493 of *NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences*, 1997. Kluwer Academic Publishers Group, Dordrecht. doi: 10.1007/978-94-011-5544-1. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1007/978-94-011-5544-1. Theory and applications.

[4] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1985. doi: 10.1007/978-1-4757-4286-2. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1007/978-1-4757-4286-2.

[5] J. O. Berger and C. Srinivasan. Generalized Bayes estimators in multivariate problems. *Ann. Statist.*, 6(4):783–801, 1978.

[6] A. R. Bernstein and F. Wattenberg. Nonstandard measure theory. In *Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967)*, pages 171–185. Holt, Rinehart and Winston, New York, 1969.

[7] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

[8] D. Blackwell and M. A. Girshick. *Theory of games and statistical decisions*. John Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London, 1954.

[9] L. D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, 42:855–903, 1971.

[10] L. D. Brown. *Fundamentals of statistical exponential families with applications in statistical decision theory*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Institute of Mathematical Statistics, Hayward, CA, 1986.

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

[12] A. D'Aristotile, P. Diaconis, and D. Freedman. On merging of probabilities. *Sankhyā Ser. A*, 50 (3):363–380, 1988. ISSN 0581-572X.

[13] R. O. Davies. Measures not approximable or not specifiable by means of balls. *Mathematika*, 18: 157–160, 1971. ISSN 0025-5793.

[14] T. S. Ferguson. *Mathematical statistics: A decision theoretic approach*. Probability and Mathematical Statistics, Vol. 1. Academic Press, New York-London, 1967.

[15] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, 2001. ISBN 0-19-857223-9.

[16] D. Heath and W. Sudderth. On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.*, 6(2):333–345, 1978.

[17] C. W. Henson. On the nonstandard representation of measures. *Trans. Amer. Math. Soc.*, 172: 437–446, 1972. ISSN 0002-9947.

[18] C. W. Henson. Analytic sets, Baire sets and the standard part map. *Canad. J. Math.*, 31(3): 663–672, 1979. ISSN 0008-414X. doi: 10.4153/CJM-1979-066-0. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.4153/CJM-1979-066-0.

[19] D. H.Fremlin. Real-valued-measurable cardinals. Version 19.9.09, https://www.essex.ac.uk/maths/people/fremlin/rvmc.pdf. Accessed 2017-02-10, 2009.

[20] W. James and C. Stein. Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 361–379. Univ. California Press, Berkeley, Calif., 1961.

[21] J. B. Kadane, M. J. Schervish, and T. Seidenfeld. Statistical implications of finitely additive probability. *Rethinking the Foundations of Statistics*, page 211, 1999.

[22] O. Kallenberg. *Foundations of modern probability*. Springer, New York, 2nd edition, 2002.

[23] H. J. Keisler. An infinitesimal approach to stochastic analysis. *Mem. Amer. Math. Soc.*, 48(297): x+184, 1984. doi: 10.1090/memo/0297. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1090/memo/0297.

[24] D. Landers and L. Rogge. Universal Loeb-measurability of sets and of the standard part map with applications. *Trans. Amer. Math. Soc.*, 304(1):229–243, 1987. doi: 10.2307/2000712. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.2307/2000712.

[25] L. LeCam. An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.*, 26: 69–81, 1955.

[26] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.

[27] 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)*, pages 144–152. Lecture Notes in Math., Vol. 369. Springer, Berlin, 1974.

[28] P. A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Trans. Amer. Math. Soc.*, 211:113–122, 1975.

[29] 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)*, pages 18–86. Holt, Rinehart and Winston, New York, 1969.

[30] N. Madras and D. Sezer. Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli*, 16(3):882–908, 2010. ISSN 1350-7265. doi: 10.3150/09-BEJ238. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.3150/09-BEJ238.

[31] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73182-9. doi: 10.1017/

CBO9780511626630. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1017/CBO9780511626630. With a prologue by Peter W. Glynn.

[32] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. in Appl. Probab.*, 25(3):487–517, 1993. ISSN 0001-8678. doi: 10.2307/1427521. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.2307/1427521.

[33] 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):518–548, 1993. ISSN 0001-8678. doi: 10.2307/1427522. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.2307/1427522.

[34] I. Molchanov. *Theory of random sets*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005. ISBN 978-185223-892-3; 1-85233-892-X.

[35] M. D. Perlman. Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *J. Multivariate Anal.*, 4:52–65, 1974. ISSN 0047-259x. doi: 10.1016/0047-259X(74)90005-0. URL http://dx.doi.org/10.1016/0047-259X(74)90005-0.

[36] D. Preiss and J. Tišer. Measures in Banach spaces are determined by their values on balls. *Mathematika*, 38(2):391–397 (1992), 1991. ISSN 0025-5793. doi: 10.1112/S0025579300006744. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1112/S0025579300006744.

[37] H. Raiffa and R. Schlaifer. *Applied statistical decision theory*. Studies in Managerial Economics. Division of Research, Graduate School of Business Administration, Harvard University, Boston, Mass., 1961.

[38] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004. ISSN 1549-5787. doi: 10.1214/154957804100000024. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1214/154957804100000024.

[39] G. O. Roberts and J. S. Rosenthal. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *Ann. Appl. Probab.*, 16(4):2123–2139, 2006. ISSN 1050-5164. doi: 10.1214/105051606000000510. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1214/105051606000000510.

[40] A. Robinson. *Non-standard analysis*. North-Holland Publishing Co., Amsterdam, 1966.

[41] 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. ISBN 0-534-05568-0.

[42] J. S. Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2006. ISBN 978-981-270-371-2; 981-270-371-3. doi: 10.1142/6300. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1142/6300.

[43] J. Sacks. Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.*, 34:751–768, 1963.

[44] L. Saloff-Coste and J. Zúñiga. Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in stochastic processes*, EMS Ser. Congr. Rep., pages 127–151. Eur. Math. Soc., Zürich, 2011. doi: 10.4171/072-1/7. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.4171/072-1/7.

[45] M. J. Schervish and T. Seidenfeld. A fair minimax theorem for two-person (zero-sum) games involving finitely additive strategies. *Rethinking the Foundations of Statistics*, pages 267–291, 1999.

[46] M. J. Schervish, T. Seidenfeld, and J. B. Kadane. The extent of nonconglomerability of finitely additive probabilities. *Z. Wahrsch. Verw. Gebiete*, 66(2):205–226, 1984. ISSN 0044-3719. doi: 10.1007/BF00531529. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1007/BF00531529.

[47] M. J. Schervish, T. Seidenfeld, and J. B. Kadane. On the equivalence of conglomerability and disintegrability for unbounded random variables. *Statistical Methods & Applications*, 23(4):501–518, 2014.

[48] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Non-conglomerability for finite-valued, finitely additive probability. *Sankhyā Ser. A*, 60(3):476–491, 1998. ISSN 0581-572X. Bayesian analysis.

[49] C. Stein. A necessary and sufficient condition for admissibility. *Ann. Math. Statist.*, 26:518–522, 1955. ISSN 0003-4851. doi: 10.1214/aoms/1177728497. URL http://dx.doi.org/10.1214/aoms/1177728497.

[50] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,*

1954–1955, vol. I, pages 197–206. University of California Press, Berkeley and Los Angeles, 1956.

[51] Ł. Stettner. Remarks on ergodic conditions for Markov processes on Polish spaces. *Bull. Polish Acad. Sci. Math.*, 42(2):103–114, 1994. ISSN 0239-7269.

[52] M. Stone. Generalized Bayes decision functions, admissibility and the exponential family. *Ann. Math. Statist.*, 38:818–822, 1967.

[53] A. Wald. Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Statistics*, 10:299–326, 1939. ISSN 0003-4851.

[54] A. Wald. Foundations of a general theory of sequential decision functions. *Econometrica*, 15: 279–313, 1947.

[55] A. Wald. An essentially complete class of admissible decision functions. *Ann. Math. Statistics*, 18:549–555, 1947.

[56] A. Wald. Statistical decision functions. *Ann. Math. Statistics*, 20:165–205, 1949.

[57] A. Wald. *Statistical Decision Functions*. John Wiley & Sons, Inc., New York, N. Y.; Chapman & Hall, Ltd., London, 1950.

[58] M. Wolff and P. A. Loeb, editors. *Nonstandard analysis for the working mathematician*, volume 510 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000. doi: 10.1007/978-94-011-4168-0. URL http://dx.doi.org.myaccess.library.utoronto.ca/10.1007/978-94-011-4168-0.