Research Announcement for:

Ergodicity of Markov Processes via Non-Standard Analysis

by

Haosui (Kevin) Duanmu, Jeffrey S. Rosenthal, and William Weiss

Departments of Statistics and Mathematics, University of Toronto

Section 1: Introduction

A time-homogeneous Markov process with a stationary probability distribution π will converge to π in an appropriate sense (i.e., will be "ergodic"), under suitable conditions (such as "irreducibility"). This phenomenon is well understood for processes in discrete time and space (see e.g. [3, 5, 13]), and for processes in continuous time and discrete space (see e.g. [5]), and for processes in discrete time and continuous space (see e.g. [8, 11]). However, for processes in continuous time and space, there are apparently no such clean results; the closest are apparently the results in [9, 10] using awkward assumptions about skeleton chains (see Section 5 below).

Meanwhile, nonstandard analysis [12] is a useful tool for providing intuitive new proofs as well as new results to all areas of mathematics, including probability and measure theory [2, 7, 1]. One of the strength of nonstandard analysis is to provide a direct passage to link discrete mathematical results to continuous mathematical results.

In the new paper [4], we apply nonstandard analysis to Markov Chain Theory. We give a nonstandard characterization of general Markov chain which allow us to view every Markov chain as a "discrete" process. The idea is to "cut" the state space and time line into hyperfinitely many pieces and we look at the transition probability between pieces at an infinitesimal time. For every standard Markov chain $\{X_t\}_{t\in\mathbb{R}}$, we can construct a hyperfinite Markov chain (that is, a Markov chain with hyperfinite state space and a hyperfinite time line) $\{X'_t\}_{t\in T}$ that inherits most of the key properties of $\{X_t\}_{t\in\mathbb{R}}$. Moreover, under moderate assumptions on $\{X_t\}_{t\in\mathbb{R}}$, we can prove the Markov Chain Ergodic theorem for all kinds of Markov processes using essentially the same argument by applying nonstandard analysis.

Section 2: Assumptions and Main Result

Definition 1. *A complete metric space X is said to satisfy the Heine-Borel condition if the closure of every bounded open set is compact.*

Consider a standard Markov process $\{X_t\}_{t\in\mathbb{R}}$ living on a σ-compact complete metric space X satisfying Heine-Borel condition. We make the following assumptions on Markov Chain $\{X_t\}_{t\in\mathbb{R}}$:

Assumption 1: Stationary distribution: There is a probability distribution π on $(X, \mathcal{B}[X])$ which is stationary, i.e. for every $A \in \mathcal{B}[X]$ and $t \in \mathbb{R}^+$ we have $\int_{x \in X} p_x^{(t)}(A)\, \pi(dx) = \pi(A)$.

We introduce the following definition before we introduce assumption 2.

Definition 2. *A Markov chain $\{X_t\}_{t\in\mathbb{R}}$ 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 $p_x^{(t)}(B) > 0$.*

Assumption 2: Productively open set irreducible: The Markov chain $\{X_t\}_{t\in\mathbb{R}}$ is productively open set irreducible, i.e. The joint Markov chain $\{X_t \times Y_t\}_{t\in\mathbb{R}}$ is open set irreducible on $X \times X$ where $\{Y_t\}_{t\in\mathbb{R}}$ is an independent identical copy of $\{X_t\}_{t\in\mathbb{R}}$.

The following lemma gives a sufficient condition for productively open set irreducible.

Lemma 1. *Let $\{X_t\}_{t\in\mathbb{R}}$ be an open set irreducible Markov process. If for any open set A and any $x \in A$, we have $P_x^{(t)}(A) > 0$ for all $t \in \mathbb{R}$. Then $\{X_t\}_{t\in\mathbb{R}}$ is productively open set irreducible.*

Assumption 3: Right continuous in time: For each fixed $x \in X$ and basic open set $A \subset X$, $p_x^{(t)}(A)$ is right continuous with left limit as a function of $t > 0$.

Assumption 4: Strong Feller: $(\forall x \in X)$ $(\forall t > 0)$ $(\forall \epsilon > 0)$ $(\exists \delta \in \mathbb{R})$ $(\forall x' \in X)$ $(|x' - x| < \delta \implies (\forall A \in \mathcal{B}[X],\ |p_{x'}^{(t)}(A) - p_x^{(t)}(A)| < \epsilon))$.

The strong Feller condition essentially says that the transition probability is continuous in state space with respect to total variation distance. For example, many diffusion processes will satisfy this property.

Assumption 5: Vanishing in Distance:
$(\forall t \in \mathbb{R}^+)(\forall \epsilon > 0)(\exists r > 0)((\forall x \in X)(\forall A \in \mathcal{B}[X])d(x, A) > r \implies P_x^{(t)}(A) < \epsilon)$.

Now we are at the place to state the main result of our paper.

Theorem 1 ([4])**.** *(Markov Chain Convergence Theorem) Let $\{X_t\}_{t\in\mathbb{R}}$ be a standard Markov chain living on a σ-compact complete metric space with Heine-Borel conditon. Assume assumptions 1-5 hold. Then for π almost surely x we have $\lim_{t\to\infty} \sup_{A\in\mathcal{B}[X]} |p^{(t)}(x, A) - \pi(A)| = 0$*

Convergence to stationary distribution in total variation distance is an important property of a Markov Process. Such convergence theorem is very well-understood for discrete time Markov Chains, but in the continuous case all of the previous results seem to require certain drift conditions or conditions on some skeleton chain. We give a proof of the continuous case with only regularity assumptions on the original chain using Nonstandard Analysis.

Section 3: Sketch of the Proof

Discrete state space discrete time Markov process is the most well-understood Markov process. We can use the classic coupling technique when proving the Markov chain Ergodic theorem for discrete Markov process. Using Nonstandard Analysis, we can essentially view every Markov process as a " discrete" Markov process.

Start with a Markov Chain $\{X_t\}_{t\in\mathbb{R}}$ satisfying the five assumptions in Markov Chain Convergence Theorem. There are 3 main steps in the proof of Markov Chain Convergence Theorem:

Step 1: Extract a Hyperfinite Markov Chain $\{X'_t\}_{t\in T}$ from $\{X_t\}_{t\in\mathbb{R}}$. A Hyperfinite Markov Chain is a Markov process with Hyperfinite state space and Hyperfinite time line. So a Hyperfinite Markov Chain behaves like a discrete time finite state space Markov Chain in many sense. The transition probability of $\{X'_t\}$ would be infinitesimally close to the transition probability of $\{X_t\}$. Moreover, $\{X'_t\}$ would inherit many key properties of $\{X_t\}$ as you will see in Theorem 4 in Section 4.

Step 2: Prove $\{X'_t\}$ converges. The proof is similar to the proof of Markov Chain Convergence Theorem for finite state space discrete time Markov process. The proof relies on the following "infinitesimal" coupling technique: For the Hyperfinite Markov Chain $\{X'_t\}$ and any two near-standard starting points i, j, they will eventually get infinitesimally close to some near-standard point i_0. $\{X'_t\}$ would "almost couple" after that.

Step 3: The convergence of $\{X'_t\}$ implies the total variation convergence of $\{X_t\}$. This is due to correspondence between the transition probability of $\{X'_t\}$ and the transition probability of $\{X_t\}$.

Thus to summarize: $\{X_t\}$ $\underrightarrow{pushup}$ $\{X'_t\}$ converges $\underrightarrow{pushdown}$ $\{X_t\}$ converges. For details, see [4].

Note that this method would work universally for all kinds of Markov processes.

Section 4: Some other Results

We also established some other results in this paper. The first two results might only be interesting to people working in nonstandard analysis.

Result 1:

Theorem 2 ([4]). *Let X be a Cech-complete then $NS({}^*X) \in {}^*\mathcal{B}[X]_L$.*

In the literature (see section 4 of [2]), people have shown this theorem for σ-compact spaces, locally compact Hausdorff spaces and complete metric spaces. Our result is a slightly generalization of the previous results.

Result 2:

Definition 3. *For a complete metric space X and any $\epsilon, r \in {}^*\mathbb{R}^+$, a hyperfinite set $S \subset {}^*X$ is* *called a (ϵ, r)-hyperfinite approximation of ${}^*X$ if the following 3 conditions hold:*

1: For each $s \in S$, there exists a $B(s) \in {}^*\mathcal{B}[X]$ with diameter no greater than ϵ containing s such that $\bigcup_{s\in S} B(s) \supset NS({}^*X)$ and the the collection of $B(s)$ for $s \in S$ are mutually disjoint.

2: For any $x \in NS({}^*X)$, ${}^*d(x, {}^*X \setminus \bigcup_{s\in S} B(s)) > r$.

3: $\bigcup_{s\in S} B(s)$ is an internal *compact subset of ${}^*X$.

Theorem 3 ([4]). *Let $(X, \mathcal{B}[X], P)$ be a Borel probability space such that X is a σ-compact complete metric space satisfying the Heine-Borel condition. Then for every $\epsilon, r \in {}^*\mathbb{R}^+$ there exists a (ϵ, r)-hyperfinite approximation S^r_ϵ of ${}^*X$. Moreover, for every (ϵ, r)-hyperfinite approximation S^r_ϵ there exists an internal probability measure P' on $(S^r_\epsilon, \mathbb{P}(S^r_\epsilon))$ such that $P(E) = \overline{P'}(st^{-1}(E) \cap S^r_\epsilon)$ where $\overline{P'}$ denotes the Loeb extension of P'*

This theorem is another Hyperfinite Representation theorem. Most known Hyperfinite representation theorems ware introduced and proved in [1]. Unlike most of the existing hyperfinite representation theorems, we restrict ourselves to σ-compact complete metric spaces in this theorem. By doing this, we are able to control the radius of each " infinitesimal piece" $B(s)$ as well as the distance between the near-standard points and those " tail points". This theorem plays an essential role in our proof of the Markov Chain Convergence Theorem.

Result 3:

Definition 4. *A Markov Chain is said to be weak Feller if:* $(\forall x \in X)$ $(\forall A \in \mathcal{B}[X])$ $(\forall t > 0)$ $(\forall \epsilon > 0)$ $(\exists \delta \in \mathbb{R})$ $(\forall x' \in X)$ $(|x' - x| < \delta \implies (|p^{(t)}_{x'}(A) - p^{(t)}_x(A)| < \epsilon))$.

It is desirable to replace the strong Feller condition by the weak Feller condition. In fact, we are able to give a "hyperfinite representation" of a standard Markov Process if we replace strong Feller by weak Feller.

Theorem 4 ([4]). *Let $\{X_t\}_{t\in\mathbb{R}}$ be a Markov chain living on a σ-compact complete metric space X satisfying Heine-Borel condition. Assume $\{X_t\}$ is weak Feller and also satisfies assumptions 1,2,3,5. Then there exists a Hyperfinite Markov chain $\{X'_t\}_{t\in T}$ such that:*

1. *The transition probability of $\{X'_t\}$ is infinitesimally close to the transition probability of $\{X_t\}$.*
2. *X_t **(productively) open set irreducible** implies that X'_t is "**(productively) open set irreducible**".*
3. *If π is a stationary distribution of X_t then hyperfinite representation measure (π') is "**almost stationary**" for X'_t.*

Section 5: Comparison with Literature

There were some similar results of the main theorem from the literature. In the literature, the following definition provides link between discrete time Markov chain and continuous time Markov chian:

Definition 5. *Given a continuous time Markov process* $\{X_t\}_{t\in\mathbb{R}}$ *and any* $m \in \mathbb{R}^+$*, a* m*-skeleton chain is a discrete Markov process living on the same state space as* $\{X_t\}$ *with one-step transition probability* $P_x^{(m)}(.)$.

As the total variation distance is non-increasing, we have that:

1: Convergence of any skeleton chain would imply the convergence of X_t.

2: Proper conditions on skeleton chain and drift conditions would reduce the problem to the discrete time Markov chain Convergence theory.

Theorems along these lines can be found in [9, 10] or in Theorem 4.1 of [6].

Generally speaking, the smaller the m is, the better the m-skeleton chain approximates X_t. Intuitively, when m is infinitesimal then the m-skeleton chain would naturally inherit all the important properties of X_t. In this way, the non-standard analysis approach in [4] reduces the general proof to the discrete-time case without making any additional assumptions on any skeleton chain.

One of the purpose of our paper [4] is to illustrates the ideas and advantages of dealing such problem with nonstandard analysis. We suspect that some of the conditions in our proof can be further weakened and one can still get the same result.

References

[1] R.M. Anderson, Star-finite representations of measure spaces. Trans. Amer. Math. Soc. **271(2)**, 667–687, 1982

[2] L.O. Arkeryd, N.J. Cutland, and C.W. Henson, Nonstandard Analysis Theory and Applications. Kluwer Academic Publishers, 1996.

[3] P. Billingsley, Probability and measure (3rd ed.). John Wiley & Sons, New York, 1995.

[4] H. Duanmu, J.S. Rosenthal, and W. Weiss, Ergodicity of Markov processes via non-standard analysis. In preparation, 2016.

[5] G.R. Grimmett and D.R. Stirzaker, Probability and random processes (2nd ed.). Oxford University Press, 1992.

[6] M. Hairer, Convergence of Markov processes. Lecture notes, University of Warwick, 2010. Available at: http://www.hairer.org/notes/Convergence.pdf

[7] H. J. Keisler, An infinitesimal approach to stochastic analysis. Mem. Amer. Math. Soc. **48(297)**, 1984.

[8] S.P. Meyn and R.L. Tweedie, Markov chains and stochastic stability. Springer-Verlag, London, 1993. Available at: http://probability.ca/MT/

[9] S.P. Meyn and R.L. Tweedie, Stability of Markovian processes II: continuous-time processes and sampled chains. Adv. Appl. Prob. **25**, 487–517, 1993.

[10] S.P. Meyn and R.L. Tweedie, Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. Adv. Appl. Prob. **25**, 518–548, 1993.

[11] G.O. Roberts and J.S. Rosenthal, General state space Markov chains and MCMC algorithms. Prob. Surv. **1**, 20–71, 2004.

[12] A. Robinson, Non-standard Analysis. Princeton University Press, 1974.

[13] J.S. Rosenthal, A First Look at Rigorous Probability Theory (2$^{\text{nd}}$ ed.). World Scientific Publishing Company, Singapore, 2006.