Von Neumann Ergodicity Theorem: An Introduction*

2024-06-17

jessekelighine.com
Jesse C. Chen 陳捷

1 Measure Preserving System and Ergodicity

Definition 1 (Measure Preserving System)**.** *A Measure Preserving System (MPS) is a quadruple $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ where $(\Omega, \mathcal{B}, \mu)$ is a measure space, and*

1. *$\mathsf{T} : \Omega \to \Omega$ is a measurable transformation,*
2. *μ is T-invariance, i.e., $\mu(\mathsf{T}^{-1}E) = \mu(E)\ \forall E \in \mathcal{B}$.*

If μ is a probability measure, then the quadruple is called a Probability Preserving System (PPS).

Remark 1. *The transformation T need not be rigid body motions when $\Omega = \mathbb{R}^n$. Consider dividing up $\mathbb{R}^n$ by grids into blocks and a transformation T that shuffles the blocks around. This is clearly a MPS when the space is equipped with Borel σ-algebra and Lebesgue measure.*

Definition 2 (Ergodic)**.** *An PPS $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ is said to be* ergodic *if $E \in \mathcal{B}$ is T-invariant, i.e., $E = \mathsf{T}^{-1}(E)$, then either $\mu(E) = 0$ or $\mu(E) = 1$.*

(a) Ergodic: well-mixed

(b) Non-ergodic: not well-mixed

Figure 1: Ergodic v.s. Non-ergodic PPS.

Remark 2. *An ergodic PPS is a system in which the only T-invariant subspaces are either negligible (measure zero) or the entire space itself. That is, an ergodic PPS is a system that is "well-mixed." In Figure 1, two systems are shown where $\Omega = \{1, \ldots, 5\}$ and transformation T is denoted by arrows. In Figure 1b, there are two non-trivial T-invariant subspaces that does not "mix" with each other; whereas in Figure 1a, every element in Ω are "mixed" with each other.*

Lemma 1. *Let $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ be an PPS, then the following are equivalent:*

1. *the PPS is ergodic,*
2. *if $E \in \mathcal{B}$ and $\mu(E \triangle \mathsf{T}^{-1}E) = 0$, then either $\mu(E) = 0$ or $\mu(E) = 1$,*
3. *if $f : \Omega \to \mathbb{R}$ is measurable and $f \circ \mathsf{T} = f$ a.e., then f is constant a.e.*

*This Introduction draws heavily from lecture note Sarig, 2023.

Proof. We show (1. $\implies$ 2.), (2. $\implies$ 3.), and (3. $\implies$ 1.) respectively.

- (1. $\implies$ 2.) Suppose there is a set $E_0 \in \mathcal{B}$ such that $E_0 = \mathsf{T}^{-1}E_0$ and $\mu(E \triangle E_0) = 0$. By ergodicity, we have $\mu(E_0) = 0$ or $\mu(\Omega - E_0) = 0$. And since $\mu(E \triangle E_0) = 0$ implies $\mu(E) = \mu(E_0)$, we have $\mu(E) = 0$ or $\mu(\Omega - E) = 0$.

  Now we construct E_0. Consider the set $E_0 = \{\omega \in \Omega : \mathsf{T}^{-k}(\omega) \in E \text{ i.o.}\}$. Clearly, E_0 is measurable and T-invariant. Also, we have $E \triangle E_0 \subseteq \bigcup_{k \geq 1} E \triangle \mathsf{T}^{-k}E$. Hence, we have

$$\begin{aligned}\mu(E \triangle E_0) &\leq \sum_{k \geq 1} \mu(E \triangle \mathsf{T}^{-k}E) \\ &\leq \sum_{k \geq 1} \sum_{j=0}^{k-1} \mu(\mathsf{T}^{-j}E \triangle \mathsf{T}^{-(j+1)}E) = \sum_{k \geq 1} k\mu(E \triangle \mathsf{T}^{-1}E)\end{aligned}$$

  where the last inequality is obtained by the fact that

$$\mu(A_1 \triangle A_2) \leq \mu(A_1 \triangle A_3) + \mu(A_3 \triangle A_2)$$

  for all $A_i \in \mathcal{B}$. Since $\mu(E \triangle \mathsf{T}^{-1}E) = 0$, we have that $\mu(E \triangle E_0) = 0$.

- (2. $\implies$ 3.) Let f be a measurable function s.t. $f \circ \mathsf{T} = f$ a.e. For any $y \in \mathbb{R}$, we have $[f > y] \triangle \mathsf{T}^{-1}[f > y] \subseteq [f \neq f \circ \mathsf{T}]$, hence

$$\mu([f > y] \triangle \mathsf{T}^{-1}[f > y]) = 0.$$

  By the assumption, either $\mu[f > y] = 0$ or $\mu[f \leq y] = 0$, i.e., either $f > y$ a.e. or $f \leq y$ a.e. Let $c \coloneqq \sup\{y \in \mathbb{R} : f > y \text{ a.e.}\}$, then $f = c$ a.e.

- (3. $\implies$ 1.) Let $E \in \mathcal{B}$ satisfies $E = \mathsf{T}^{-1}(E)$. Consider $f = \mathbf{1}_E$. Since $f \circ \mathsf{T} = f$, f is constant a.e., we have $f = 0$ a.e. or $f = 1$ a.e., implying that either $\mu(E) = 0$ or $\mu(E) = 1$. #

Remark 3. *The third characterization is quite interesting and intuitive. Consider again Figure 1b, where the PPS is equipped with probability space* $\Omega = \{1, ..., 5\}$, $\mathcal{B} = 2^{\Omega}$, *and uniform* μ. *We can define the function* f *as follows:*

$$f(\omega) = \begin{cases} 0 & \text{if } \omega \in \{1, 2, 3\}, \\ 1 & \text{otherwise.} \end{cases}$$

It is clear that $f \circ \mathsf{T} = f$, *but* f *is not constant on* Ω. *This is achievable since non-ergodicity means there are non-trivial* T*-invariant subspaces, and we can simply define* f *to be constant on each subspace. However, in Figure 1a, since the system is "well-mixed," this trick is not possible.*

Definition 3 (Strong Mixing)**.** *An PPS* $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ *is called* strong mixing *if for any* $E, F \in \mathcal{B}$ *we have*

$$\mu(E \cap \mathsf{T}^{-n}F) \to \mu(E)\mu(F) \quad \text{as} \quad n \to \infty.$$

Remark 4. *The fact that strong mixing is defined using the inverse map of* T *is not merely due to measure-theoretic technicalities. The interpretation is "no matter what obscure events* F *one chooses, it could have been from all over in the system entirely randomly, not just some specific part." Thus, for any other event* E*, the "origins" of* F *must be as if independent of* E.

Lemma 2. *Strong mixing implies ergodicity.*

Proof. Suppose $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ is strong mixing. Let E be such that $E = \mathsf{T}^{-1}E$, then we have

$$\mu(E) = \mu(E \cap \mathsf{T}^{-n}E) \to \mu(E)^2 \quad \text{as} \quad n \to \infty.$$

Clearly, $\mu(E) = \mu(E)^2$ implies $\mu(E)$ is either 0 or 1. #

2 Ergodicity Theorem

Theorem 1 (von Neumann's Ergodicity)**.** *Let $\mathcal{H}$ be a Hilbert space. Let $\mathsf{U} : \mathcal{H} \to \mathcal{H}$ be a unitary operator. Let $\mathcal{I} = \{f \in \mathcal{H} : \mathsf{U}f = f\}$ denote the* U*-invariant subspace. Let $\mathsf{P} : \mathcal{H} \to \mathcal{I}$ be the orthogonal projection onto $\mathcal{I}$. Then, for any $f \in \mathcal{H}$, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i f \to \mathsf{P}f \quad \text{as} \quad n \to \infty \tag{1}$$

in the norm induced by the inner product on $\mathcal{H}$.

Proof. Clearly, Equation 1 holds when $f \in \mathcal{I}$. Let $\mathcal{J} \coloneqq \{g - \mathsf{U}g : g \in \mathcal{H}\}$. Suppose $f \in \mathcal{J}$, we have

$$\langle f, h\rangle = \langle g, h\rangle - \langle \mathsf{U}g, h\rangle = \langle g, h\rangle - \langle g, \mathsf{U}h\rangle = 0 \quad \forall h \in \mathcal{I}.$$

Hence, $\mathsf{P}f = 0$. Furthermore, we have

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i f \right\|_2 = \frac{1}{n}\left\| \mathsf{U}g - \mathsf{U}^{n+1}g \right\|_2 \le \frac{1}{n}\|2g\|_2 \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore, Equation 1 holds for $\mathcal{J}$.

We claim that Equation 1 also holds for the closure of $\mathcal{J}$, denoted by $\bar{\mathcal{J}}$. Suppose $f \in \bar{\mathcal{J}}$, then $\forall \varepsilon > 0\ \exists g \in \mathcal{J}$ s.t. $\|f - g\|_2 < \varepsilon$. Choose N such that $\left\| \frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i g \right\|_2 < \varepsilon$ $\forall n > N$. Then, we have

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i f \right\|_2 \le \left\| \frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i (f-g) \right\|_2 + \left\| \frac{1}{n}\sum_{i=1}^{n} \mathsf{U}^i g \right\|_2 \le 2\varepsilon.$$

Thus, the claim holds.

Lastly, we claim that $\bar{\mathcal{J}}^{\perp} = \mathcal{I}$. Suppose $f \perp \bar{\mathcal{J}}$, we have that $\forall g \in \mathcal{H}$,

$$0 = \langle f, g - \mathsf{U}g\rangle = \langle f, g\rangle - \langle f, \mathsf{U}g\rangle \implies \langle f, g\rangle = \langle \mathsf{U}^* f, g\rangle.$$

This implies that $f = \mathsf{U}f$ a.e., i.e., $f \in \mathcal{I}$. Conversely, we have shown that if $f \in \mathcal{J}$, then $f \perp \mathcal{I}$. Hence, the claim holds. And since $\mathcal{H} = \bar{\mathcal{J}} \oplus \mathcal{I}$, Equation 1 holds for all $f \in \mathcal{H}$. #

Remark 5. *If we think about Theorem 1 in linear algebra terms, it lends itself to an intuitive understanding. In Figure 2, we consider $\mathbb{R}^3$ space with* U *being a rotation along the z-axis. Clearly, a rotation is an unitary transformation, and its corresponding invariant subspace is exactly the z-axis. Consider an arbitrary vector f and its transformations $\mathsf{U}^i f$. As $\mathsf{U}^i f$ rotates along the z-axis, it is clear that their "average" converges to $\mathsf{P}f$, the orthogonal projection of f onto the z-axis. One can also easily see that vectors of the form $f - \mathsf{U}f$ is orthogonal to the z-axis.*

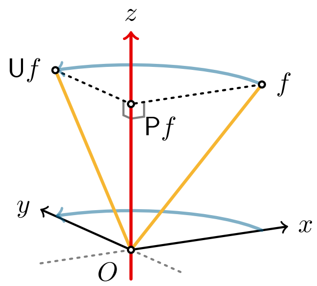


Figure 2: Visualization of Theorem 1.

Corollary 1.1. *Let* $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ *be an PPS. If* $f \in L^2(\Omega)$*, then*

$$\frac{1}{n}\sum_{i=1}^{n} f \circ \mathsf{T}^i \xrightarrow{L^2} \bar{f} \quad \text{as} \quad n \to \infty$$

where $\bar{f} \in L^2(\Omega)$ *is* T*-invariant. Furthermore, if* $(\Omega, \mathcal{B}, \mu, \mathsf{T})$ *is ergodic, then*

$$\bar{f} = \int_{\Omega} f \, d\mu.$$

Proof. Note that $L^2(\Omega)$ is a Hilbert space. Let $\mathsf{U}f := f \circ \mathsf{T}$. Since $\|f \circ \mathsf{T}\|_2 = \|f\|_2$ for $f \in L^2(\Omega)$, U is unitary. Hence, the statement is a direct implication of Theorem 1. Furthermore, if the PPS is ergodic, then by Lemma 1, the T-invariant subspace $\mathcal{I}$ contains only constant functions. Therefore, the orthogonal projection of f onto $\mathcal{I}$ is its expectation. #

Remark 6. *The sum still converges without ergodicity, but this is not very useful, since we do not know the form of* $\bar{f}$*. With ergodicity, we know* $\bar{f}$ *is a constant function. Hence, we can simply pick any starting point* $\omega \in \Omega$ *and compute* $f \circ \mathsf{T}^i(\omega)$ *along the way to obtain the same constant* $\bar{f}$*, without worrying about the entire functional space.*

3 Markov Chains under the Language of Ergodicity

Now we want to rephrase what we already understand about Markov chains under the language of von Neumann's Ergodicity Theorem.

Definition 4 (Subshift of Finite Type)**.** *Let* $\mathcal{S}$ *with be a finite set of states. Let* $\mathsf{A} = (\mathsf{A}_{ij})_{\mathcal{S}\times\mathcal{S}}$ *be a adjacency matrix with* $\mathsf{A}_{ij} \in \{0, 1\}$ *and without rows or columns be entirely zero. Let* Ω_{A} *be defined as the space of possible sequences*

$$\Omega_{\mathsf{A}} := \{\omega = (\omega_0, \omega_1, \ldots) : \mathsf{A}_{\omega_i,\omega_{i+1}} = 1 \ \forall i\}.$$

A Subshift of Finite Type (SFT) with states $\mathcal{S}$ *with adjacency matrix* A *is the triple* $(\Omega_{\mathsf{A}}, d, \mathsf{T})$ *where* $d(\omega, \omega') = 2^{-\min\{k:\omega_k \neq \omega'_k\}}$ *is a metric and* $\mathsf{T}(\omega_0, \omega_1, \ldots) = (\omega_1, \omega_2, \ldots)$ *is a shift transformation.*

Remark 7. *The topology generated by the metric $d(\cdot,\cdot)$ is equivalent to the topology generated by sets of the form*

$$[s_0, ..., s_{n-1}] \coloneqq \{\omega \in \Omega_{\mathsf{A}} : \omega_i = s_i \ \forall i \in \{0, ..., n-1\}\}.$$

This form are sometimes referred to as "cylinders." This is the product topology on $\mathcal{S}^{\mathbb{N}}$ when $\mathcal{S}$ is given the discrete topology.

Remark 8. *The space Ω is simply all possible sequencing of $\mathcal{S}$ that is permitted by A. Consider again Figure 1, now with the arrows denoting the adjacency matrix A, then Figure 1a* and *Figure 1b generates different Ω_{A}'s. What we want next is a way to say "how likely is any given sequence realized/observed."*

Definition 5 (Markov Measure)**.** *Given a transition matrix P, i.e., a matrix with each row consisting weights summing to one, and a probability vector π, we construct a measure μ by letting*

$$\mu[s_0, ..., s_{n-1}] \coloneqq \pi_{s_0} \mathsf{P}_{s_0 s_1} \cdots \mathsf{P}_{s_{n-2} s_{n-1}}.$$

Remark 9. *The definition of Markov measure extends to a unique probability measure on Ω_{A} with product topology via Carathéodory extension theorem.*

Lemma 3. *μ is T-invariant iff π is stationary w.r.t. P.*

Proof. Note that μ is T-invariant iff $\mu[\cdot, \mathbf{s}] = \mu[\mathbf{s}] \ \forall \mathbf{s} = (s_0, ..., s_{n-1})$. That is,

$$\sum_{t \in \mathcal{S}} \pi_t \mathsf{P}_{t s_0} \mathsf{P}_{s_0 s_1} \cdots \mathsf{P}_{s_{n-2} s_{n-1}} = \pi_{s_0} \mathsf{P}_{s_0 s_1} \cdots \mathsf{P}_{s_{n-2} s_{n-1}} \quad \forall \mathbf{s}$$
$$\iff \sum_{t \in \mathcal{S}} \pi_t \mathsf{P}_{t s_0} = \pi_{s_0} \quad \forall s_0 \in \mathcal{S}. \qquad \#$$

Remark 10. *With Lemma 3, we established that the system corresponding to the SFT is a measure preserving system as long as the Markov measure is constructed with a stationary distribution. It remains to ask the question: When is such system ergodic? When is such system strong mixing? From the intuition established previously, we know that...*

- Ergodic *means that the system is "well-mixed," i.e., all the states should be able to be reached from any other state. This corresponds to the idea of* irreducibility.
- Strong mixing *means that the "origin" of any set could be all over the system entirely randomly, i.e., we shouldn't observe a set that follows some path. This corresponds to the idea of* aperiodicity.

It would turn out that our intuitions are spot on in describing what kinds of SFT are ergodic and strong mixing.

Definition 6 (Irreducible)**.** *A transition matrix P is called* irreducible *if $\forall a, b \in \mathcal{S}$ $\exists s_1, ..., s_{n-1} \in \mathcal{S}$ s.t. $\mathsf{P}_{a,s_1} \cdots \mathsf{P}_{s_{n-1},b} > 0$. We denote this as $a \xrightarrow{n} b$.*

Lemma 4. *If P is irreducible, then $\gcd\{n \in \mathbb{N} : a \xrightarrow{n} a\}$ is independent of a.*

Proof. Suppose $p_a = \gcd\{n : a \xrightarrow{n} a\}$ and $p_b = \gcd\{n : b \xrightarrow{n} b\}$ for $a, b \in \mathcal{S}$. Since P is irreducible, we have $a \xrightarrow{m_1} b \xrightarrow{m_b} b \xrightarrow{m_2} a$ for some $m_1, m_2 \in \mathbb{N}$ and $m_b \in \{n : b \xrightarrow{n} b\}$. Since $p_a \,|\, m_1 + m_b + m_2$ and $p_b \,|\, m_b$, we must have $p_a \,|\, p_b$. Similarly, we have $p_b \,|\, p_a$ by swapping a and b. Therefore, $p_a = p_b$. #

Definition 7 (Period)**.** *The* period *of an* P *is* $\gcd\{n : a \xrightarrow{n} a\}$. *An irreducible* P *is* aperiodic *if* $\gcd\{n : a \xrightarrow{n} a\} = 1$.

Remark 11. *Definition 7 is well-defined since Lemma 4 guarantees that as long as* P *is irreducible, then the period is independent of state.*

Theorem 2 (Ergodicity for Markov Chains)**.** *Let* P *be a transition matrix and let* P^n_{ab} *denote the* (a, b)*-th element of* P^n. *Let* π *be a stationary distribution of* P. *Then,*

1. *if* P *is irreducible, then as* $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{P}^n_{ab} \to \pi_b \quad \forall a, b \in \mathcal{S}.$$

2. *if* P *is irreducible and aperiodic, then as* $n \to \infty$,

$$\mathsf{P}^n_{ab} \to \pi_b \quad \forall a, b \in \mathcal{S}.$$

Proof. Omitted. See, e.g., Chapter 1.5 in Sarig, 2023 or Chapter 5.6 in Durrett, 2019 for a proof. #

Corollary 2.1. *Let* $(\Omega_{\mathsf{A}}, \mathcal{B}, \mu, \mathsf{T})$ *be the PPS corresponding to the SFT* $(\Omega_{\mathsf{A}}, d, \mathsf{T})$ *with Markov measure* μ *under transition matrix* P. *If* P *is irreducible, then the PPS ergodic. Furthermore, if* P *is also aperiodic, then the PPS is strong mixing.*

Proof. First, note that for all cylinders $[\boldsymbol{a}] = [a_0, ..., a_{n_a - 1}]$ and $[\boldsymbol{b}] = [b_0, ..., b_{n_b - 1}]$ and $k > n_a$, we have

$$\begin{aligned}
\mu([\boldsymbol{a}] \cap \mathsf{T}^{-k}[\boldsymbol{b}]) &= \mu\left(\bigsqcup_{\boldsymbol{c} \in \mathcal{C}_{k-n_a}} [\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}]\right) \\
&= \mu[\boldsymbol{a}]\left(\sum_{\boldsymbol{c} \in \mathcal{C}_{k-n_a}} \mathsf{P}_{a_{n_a-1}, c_0} \cdots \mathsf{P}_{c_{k-n_a-1}, b_0}\right)\frac{\mu[\boldsymbol{b}]}{\pi_{b_0}} \\
&= \mu[\boldsymbol{a}]\mu[\boldsymbol{b}]\frac{\mathsf{P}^{k-n_a}_{a_{n_a-1}, b_0}}{\pi_{b_0}} \qquad (2)
\end{aligned}$$

where $\mathcal{C}_\ell := \{\boldsymbol{c} = (c_0, ..., c_{\ell-1}) : [\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{b}] \neq \varnothing\}$. By Theorem 2, we have that

$$\frac{1}{n}\sum_{k=n_a+1}^{n} \mu([\boldsymbol{a}] \cap \mathsf{T}^{-k}[\boldsymbol{b}]) = \mu[\boldsymbol{a}]\mu[\boldsymbol{b}]\frac{1}{\pi_{b_0}}\left(\frac{1}{n}\sum_{k=n_a+1}^{n} \mathsf{P}^{k-n_a}_{a_{n_a-1}, b_0}\right) \to \mu[\boldsymbol{a}]\mu[\boldsymbol{b}].$$

Now suppose that P is irreducible. Let $E \in \mathcal{B}$ be invariant, then $\exists [\boldsymbol{a}_1], ..., [\boldsymbol{a}_m]$ s.t. $\mu(E \triangle \textstyle\bigplus_{j=1}^m [\boldsymbol{a}_j]) < \varepsilon$. Since the collection of cylinders forms a semi-algebra, such $[\boldsymbol{a}_1], ..., [\boldsymbol{a}_m]$ always exists. Then,

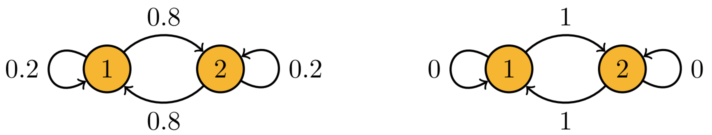
$$\mu(E) = \mu(E \cap \mathsf{T}^{-k} E) = \sum_{i=1}^m \sum_{j=1}^m \mu([\boldsymbol{a}_i] \cap \mathsf{T}^{-k}[\boldsymbol{a}_j]) \pm 2\varepsilon.$$

Taking average over k, by the fact we mentioned above, we have

$$\begin{aligned} \mu(E) &= \sum_{i=1}^m \sum_{j=1}^m \left(\frac{1}{n} \sum_{k}^{n} \mu([\boldsymbol{a}_i] \cap \mathsf{T}^{-k}[\boldsymbol{a}_j]) \right) \pm 2\varepsilon \\ &\to \sum_{i=1}^m \sum_{j=1}^m \mu[\boldsymbol{a}_i]\mu[\boldsymbol{a}_j] \pm 2\varepsilon = \left(\sum_{i=1}^m \mu[\boldsymbol{a}_j] \right)^2 \pm 2\varepsilon = (\mu(E) \pm \varepsilon)^2 \pm 2\varepsilon \end{aligned}$$

as $n \to \infty$. Therefore, take $\varepsilon \downarrow 0$ and we have that $\mu(E) = \mu(E)^2$, which implies either $\mu(E) = 0$ or $\mu(E) = 1$. Hence, we have ergodicity.

Now suppose further that P is aperiodic. By Theorem 2 and (2), we have $\mu([\boldsymbol{a}] \cap \mathsf{T}^{-k}[\boldsymbol{b}]) \to \mu[\boldsymbol{a}]\mu[\boldsymbol{b}]$ as $k \to \infty$. By a similar argument with approximation using cylinders, we obtain the desired result. #



(a) Aperiodic (b) Periodic

Figure 3: Markov Chains under different transition matrix P.

Remark 12. *Figure 3 shows two Markov chains with same adjacency A but different transition P. One can easily check both chains represented in Figure 3a and Figure 3b share the same invariant distribution $\pi = (0.5, 0.5)$ However, notice that Figure 3b has period 2. Hence, P_{11}^n converges to 0.5 in Figure 3a but fails to converge in Figure 3b.*

Remark 13. *Corollary 2.1 can be extended to an if-and-only-if statement. The reverse statement can be proven using techniques similar to the proof of Corollary 2.1.*

Acronyms

MPS Measure Preserving System. 1
PPS Probability Preserving System. 1, 2, 4, 6
SFT Subshift of Finite Type. 4–6

References

Durrett, R. (2019). Probability: Theory and examples. https://services.math.duke.edu/~rtd/PTE/pte.html

Sarig, O. (2023). Lecture notes on ergodic theory. https://www.weizmann.ac.il/math/sarigo