



APPROXIMATION WITH ERGODIC PROCESSES AND TESTABILITY

ISAAC LOH ,* *UNC Wilmington*

Abstract

We show that stationary time series can be uniformly approximated over all finite time intervals by mixing, non-ergodic, non-mean-ergodic, and periodic processes, and by codings of aperiodic processes. A corollary is that the ergodic hypothesis—that time averages will converge to their statistical counterparts—and several adjacent hypotheses are not testable in the non-parametric case. Further Baire category implications are also explored.

Keywords: Mixing; testability; ergodicity

2020 Mathematics Subject Classification: Primary 37A50
Secondary 62G10

1. Introduction

This paper establishes uniform approximation results which state that any time series taking on values in a Polish space can be approximated in a strong sense by stationary ergodic (in fact, mixing) time series. This theorem is proved by applying results on the uniform approximation of measure-preserving transformations to their time series analogues. The approximation results also hold very generally with non-ergodic processes, mean-ergodic processes, and periodic series, as well as for codings of arbitrary aperiodic time series.

An immediate corollary of our approximation result is that, in our non-parametric setting, the power of any test for ergodicity (or alternately mixing, non-ergodicity, mean-ergodicity, non-mean-ergodicity, periodicity) cannot exceed its size, despite the number of time periods or the number of observations available. In other words, the ergodic hypothesis, and several related hypotheses, are not testable. This finding contrasts with the tests for ergodicity which are proposed in [11, 13, 14] for series that are Markovian in nature. It also contrasts with consistent estimation schemes that exist when the space of time series is restricted appropriately (e.g., [34] considers the space of B -processes). Recent applications of non-parametric tests for ergodicity occur in settings arising from economics and the physical sciences [19, 21, 32, 36, 45].

Our approximation results complement some well-known results for stationary processes. [20, 31] show the existence of stationary processes with a finite state space arising from measure-preserving, ergodic, and aperiodic transformations that have a certain prescribed marginal distribution. Similarly, [3, 4, 30] study the problem of mapping (coding) a stationary process onto another process with a given marginal distribution when the target stochastic

Received 21 January 2022; accepted 26 October 2023.

* Postal address: Department of Economics and Finance, UNC Wilmington, 601 South College Road, Wilmington NC 28403. Email: lohi@uncw.edu

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust.

process takes on values in a countable set. Our findings on approximations and codings of aperiodic processes especially parallel those of [29], which shows that marginal distributions for stationary processes on a finite state space can be matched with periodic measures and closely approximated with ergodic measures.

Approximation and Baire category results have generated considerable interest in ergodic theory (cf. [9, 10, 17] for surveys), and have been extended to the study of stochastic processes. [35] shows that the set of ergodic measures for stationary time series is a generic set in the set of stationary measures endowed with the weak topology, whereas the set of mixing measures is meagre. [17] extends these results to general transformations, and [9] to a range of measure-theoretic properties on Polish spaces. We apply our results to provide a similar treatment in a metric space of stationary time series. In this space, we show that weak mixing is the strongest of a range of mixing conditions which hold for a topologically large set of time series.

2. Main result

Fix a Polish space \mathcal{X} . Let $\mathcal{X}^{\mathbb{Z}}$ denote $\prod_{t \in \mathbb{Z}} \mathcal{X}$, and let $T : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ be the left shift map. We begin by defining some properties of time series (see, e.g., [28]). An \mathcal{X} -valued time series $X = (X_t(\omega))_{t \in \mathbb{Z}} : (\Omega, \mathfrak{F}, \mathbb{P}) \rightarrow \mathcal{X}^{\mathbb{Z}}$ is *stationary* if its finite-dimensional distributions are shift invariant, i.e. for every cylinder set $A \subset \mathcal{X}^{\mathbb{Z}}$, $\mathbb{P}(T^t X \in A)$ is invariant with respect to t . A time series is *ergodic* if, for all cylinder sets $A \subset \mathcal{X}^{\mathbb{Z}}$,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t \mathbf{1}_A(T^j X) \stackrel{\text{a.s.}}{=} \mathbb{P}(X \in A).$$

We follow the convention of dynamical systems and say that a stationary process is *weakly mixing* if, for all Borel measurable $A, B \subset \mathcal{X}^{\mathbb{Z}}$,

$$\lim_{t \rightarrow \infty, t \notin I_0} \mathbb{E}[\mathbf{1}_A(T^t X) \mathbf{1}_B(X)] = \mathbb{P}(A) \mathbb{P}(B), \quad (1)$$

where I_0 is some zero-density subset of \mathbb{N} (recall that $I_0 \subset \mathbb{N}$ is said to have zero density if $\lim_{t \rightarrow \infty} 1/t \sum_{j=1}^t \mathbf{1}_{j \in I_0} = 0$). Weak mixing implies ergodicity. A process is *mixing* if it satisfies (1) with $I_0 = \emptyset$. This version of mixing is far weaker than some common notions of mixing in time series, such as α -, β -, φ -, and ψ -mixing.

In the particular case where $\mathcal{X} = \mathbb{R}$, we call a stationary time series *mean ergodic* if

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t X_j \stackrel{\text{a.s.}}{=} \mathbb{E}[X_0],$$

provided that the integral on the right exists. X is *aperiodic* if the probability that $\dots, X_{-1}, X_0, X_1, \dots$ is a periodic sequence is zero. A time series is *periodic*, with period p , if $X_{t+p} \stackrel{\text{a.s.}}{=} X_t$ for all $t \in \mathbb{Z}$. Recall that a probability space is called a *standard* probability space if it is isomorphic to a regular probability measure on the unit interval and a countable number of atoms (c.f. [26, §2.4] or [7, §9.4]). We say that a random variable is non-degenerate if it does not almost surely (a.s.) equal one fixed constant.

If X is supported on a standard probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, then we construct sequences (X^k) approximating X on an extended sample space $(\Omega \times \Omega', \mathfrak{F} \otimes \mathfrak{F}', \mathbb{P} \times \mathbb{P}')$, where $(\Omega', \mathfrak{F}', \mathbb{P}')$ is again standard. The introduction of additional randomness to the sample

space via Ω' is a necessary step in even approximating periodic processes with aperiodic ones (Remark 1). Whenever this is the case, we regard X as a random variable defined on $\Omega \times \Omega'$ which simply does not depend on its second coordinate. In particular, we make the identification

$$X \sim X \circ \pi_1 : \Omega \times \Omega' \rightarrow \mathcal{X}^{\mathbb{Z}}, \tag{2}$$

where π_1 is the projection onto the first coordinate, and with abuse of notation refer to the product measure $\mathbb{P} \times \mathbb{P}'$ by \mathbb{P} . This technical imposition ensures that \mathbb{P} indeed describes the joint distribution of X and (X^k) , and affords some economy of notation. The expression in (2) is a formal way of adopting the convention that probabilistic concepts must be preserved under extension of the underlying sample space (cf. [43, §1]). All of our results are equally valid if one regards the right side of (2) as merely a distributional copy of X defined on $\Omega \times \Omega'$.

Our first theorem shows that X can be approximated over all finite time intervals by aperiodic time series having mixing and non-mixing characteristics.

Theorem 1. *Let \mathcal{X} be a Polish space satisfying $|\mathcal{X}| \geq 2$. Then, for any stationary \mathcal{X} -valued time series X on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, and any standard non-atomic probability space $(\Omega', \mathfrak{F}', \mathbb{P}')$, there is a sequence (X^k) of stationary, aperiodic, and mixing \mathcal{X} -valued time series on $(\Omega \times \Omega', \mathfrak{F} \otimes \mathfrak{F}', \mathbb{P} \times \mathbb{P}')$ such that, for all $t \in \mathbb{N}$,*

$$\lim_{k \rightarrow \infty} \mathbb{P}(X_{-t} = X_{-t}^k, \dots, X_t = X_t^k) = 1, \tag{3}$$

and, moreover, if X is non-degenerate, $X_0 \stackrel{d}{=} X_0^k$ for all $k \in \mathbb{N}$. The same is true if, instead of mixing, (X^k) is specified to be non-ergodic. If X is in addition an integrable \mathbb{R} -valued time series, the same is true if, instead of mixing, (X^k) is specified to be mean-ergodic or non-mean-ergodic.

The proof of Theorem 1 proceeds in two steps. The first step identifies each time series with a measure-preserving (m.p.) transformation on a standard probability space (Proposition 3). The second step approximates these transformations in a uniform sense with, say, mixing transformations, and converts the approximating transformations back into time series (Proposition 4). Proofs of the main results are given in Section 4.

The proof of Theorem 1 and its implications for non-testability also resemble the findings of [1], which applies the cutting-and-stacking method of ergodic theory [16] to show that it is impossible to guarantee consistent estimation of the one-dimensional marginal densities for a stationary ergodic process. There is related literature on learnability under stationary ergodicity which makes use of ergodic approximations to establish impossibility results (see [24, 39] and references therein), although the discernability of processes fulfilling various mixing properties from non-ergodic processes has not been well studied. It would be interesting to see if the dynamical system approximation results of [16], which are quite strong and used in the proof of Theorem 1, have any applications in this direction.

The conclusion of Theorem 1 fails if the probability space Ω' is not introduced to the domain of the approximating time series via, e.g., (2).

Remark 1. We can show that if $X_0 : \Omega \rightarrow \mathcal{X}$ is an injection, and X fails to be aperiodic, the approximation in Theorem 1 cannot occur even with aperiodic time series without extending the sample space Ω . Indeed, let X be such a series, so that for some period $p \geq 0$, $\mathbb{P}(X_0 = X_{tp}) > 0$ for all $t \in \mathbb{Z}$.

Let $X' : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ be a stationary and aperiodic time series for which $X'_0 \stackrel{d}{=} X_0$, as in Theorem 1. Let $\Omega_1 = \{\omega : X'_0 = X_0 = X_p = X'_p\}$ and $\Omega_2 = \Omega \setminus \Omega_1$. By assumption, $\mathbb{P}(X'_0 \in X_0(\Omega_2)) = \mathbb{P}(X_0 \in X_0(\Omega_2)) = \mathbb{P}(\Omega_2)$. Moreover, $X'_0(\omega) \in X_0(\Omega_2)$ only if $\omega \in \Omega_2$. Hence, $(X'_0)^{-1}(X_0(\Omega_2)) = \Omega_2 \pmod{\mathbb{P}}$, and $(X'_0)^{-1}(X_0(\Omega_1)) = \Omega_1 \pmod{\mathbb{P}}$. Arguing similarly for X'_p , it follows that X'_p is in $X_0(\Omega_1)$ if and only if X'_0 is.

Suppose that $\mathbb{P}(\Omega_1) > 0$, which must eventually happen for X'_k if (3) holds. From the above, we have $\mathbb{P}(X'_p \in X_0(\Omega_1) \mid X'_0 \in X_0(\Omega_1)) = \mathbb{P}(X'_p(\omega) \in X_0(\Omega_1) \mid \omega \in \Omega_1) = 1$, and $\mathbb{P}(X'_p \in X_0(\Omega_2) \mid X'_0 \in X_0(\Omega_2)) = \mathbb{P}(X'_p(\omega) \in X_0(\Omega_2) \mid \omega \in \Omega_2) = 1$. Induction and stationarity then imply that $\mathbb{P}(X'_{tp} \in X_0(\Omega_1) \mid X'_0 \in X_0(\Omega_1)) = 1$ for all $t \in \mathbb{Z}$. In fact, because $X'_p = X'_0$ on $X_0(\Omega_1)$, the inductive argument also implies that $\mathbb{P}(X'_{tp} = X'_0 \text{ for all } t \in \mathbb{Z} \mid X'_0 \in X_0(\Omega_1)) = 1$. Thus, X' is not even aperiodic.

The distributional result $X_0^k \stackrel{d}{=} X_0$ of Theorem 1 holds only if X is non-degenerate. To see why this is the case, note that if X is degenerate and equals some $x \in \mathcal{X}^{\mathbb{Z}}$ almost surely, then X_0 is also degenerate and equals x_0 , which is the zeroth coordinate of x . Then, if $X_0^k \stackrel{d}{=} X_0$, $X_0^k \stackrel{\text{a.s.}}{=} x_0$, and stationarity requires $X_t^k = x_0$ for all t , so X^k cannot be aperiodic. It is also non-negligible to assume that Ω' is non-atomic. For instance, in the worst case, Ω' is just a single point mass, and the preceding example shows that Theorem 1 no longer holds. If Ω' consists of a finite number of atoms $\omega'_1, \dots, \omega'_M$, we can take $\Omega_1 = \{\omega \in \Omega : X'_0(\omega, \omega') = X_0(\omega) = X'_p(\omega, \omega') = X_p(\omega) \text{ for all } \omega' \in \Omega'\}$, and argue similarly to the example above.

The problem of coding a stationary stochastic processes to achieve certain marginal distributions has received attention in information theory (cf. [3, 4, 20, 30, 31]). While Theorem 1 does not exactly replicate the marginal distributions of X , it does build an approximating sequence for X itself, and not just its marginal distributions. It also has the additional ability to construct the approximating series as mixing, non-ergodic, or non-mean ergodic processes over arbitrary Polish spaces.

A slight variation on the proof of Theorem 1 establishes a similar approximation result for codings of time series taking on values in any Polish space. For a given time series $Y : \Omega \rightarrow \mathcal{Y}^{\mathbb{Z}}$, where \mathcal{Y} is a Polish space, a coding X of Y is a time series $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ satisfying

$$X_t = \tilde{\chi}(S^t Y), \tag{4}$$

where $S : \mathcal{Y}^{\mathbb{Z}} \rightarrow \mathcal{Y}^{\mathbb{Z}}$ is the left shift map, and $\tilde{\chi} : \mathcal{Y}^{\mathbb{Z}} \rightarrow \mathcal{X}$ is called the *time-zero coder* [40]. A *finite coding* is a special case where there is some window length $w < \infty$ such that, if y and y' agree on their $-w$ th through w th coordinates, then $\tilde{\chi}(y) = \tilde{\chi}(y')$.

In the proofs of Theorems 2 and 3, we continue to identify X with the time series supported on $\Omega \times \Omega'$, where Ω' is a standard non-atomic probability space, using (2).

Theorem 2. *Let \mathcal{X} and \mathcal{Y} be Polish spaces and X any \mathcal{X} -valued time series on Ω . Let Q be the law of any stationary and aperiodic \mathcal{Y} -valued time series, and Ω' be a standard non-atomic probability space. Then there is a sequence (Y^k) of time series $Y^k : \Omega \times \Omega' \rightarrow \mathcal{Y}^{\mathbb{Z}}$ with law Q and codings X^k of the Y^k such that (3) holds for all $t \in \mathbb{N}$, and $X_0 \stackrel{d}{=} X_0^k$ for all $k \in \mathbb{N}$.*

As properties such as mixing and ergodicity are preserved under coding, Theorem 2 implies the mixing result of Theorem 1 as long as the approximating time series are not required to be aperiodic. Theorem 2 makes use of a uniform approximation result for m.p. transformations. If the assumption of stationarity is also dropped, similar approximation results for non-singular transformations may be applied in a similar fashion [16, §7].

Aperiodicity of Q is a non-negligible assumption in Theorem 2, and was also used in [3] (consider, for instance, the worst case where Q is deterministic). However, it may also be dropped when the approximating series are themselves periodic.

Theorem 3. *Let \mathcal{X} be Polish, X any \mathcal{X} -valued time series on Ω , and Ω' a standard non-atomic probability space. Then there is a sequence of stationary and periodic time series $X^k : \Omega \times \Omega' \rightarrow \mathcal{X}^{\mathbb{Z}}$ such that (3) holds for all $t \in \mathbb{N}$, and $X_0 \stackrel{d}{=} X_0^k$ for all $k \in \mathbb{N}$.*

2.1. Relation to joinings and entropy

Theorems 1–3 produce examples of jointly distributed stochastic processes (X, X^k) on $\mathcal{X}^{\mathbb{Z}} \times \mathcal{X}^{\mathbb{Z}}$, where X^k approximates X arbitrarily closely on its finite-dimensional marginals. The joint distribution is a coupling of the distribution of X and that of X^k . If this distribution were itself stationary, it would be termed a joining [40]. Such joinings have been studied extensively in optimal transport, where they define the well-known \bar{d} -distance [33] and generalizations yielding optimal joining costs with different metrics [18, 38].

The couplings provided by Theorems 1 and 2 are not in general stationary, nor indeed can they always be stationary for ergodic (mixing) approximations of non-ergodic (non-mixing) time series. In the simple case where \mathcal{X} is a finite alphabet, this follows because the optimal joining cost between ergodic and non-ergodic processes must be bounded away from zero. Suppose that the joint distribution of (X^k, X) was a joining for all k . Then, letting (X^k) be ergodic (mixing) in the statement of Theorem 1 or 2, we must have $\bar{d}(\mu_k, \mu) \rightarrow 0$, where μ_k is the law of X^k and μ is the law of X (see, e.g., [40, Theorem I.9.7]). [40, Theorems I.9.15 and I.9.17] would then imply that X was ergodic (mixing), which is impossible, because X is arbitrary in Theorems 1 and 2. This argument can be extended to non-discrete sample paths via the following simple lemma.

Lemma 1. *For any non-ergodic (non-mixing) \mathcal{Y} -valued stationary time series Y , there exists a finite \mathcal{X} and \mathcal{X} -valued time-zero coder $\tilde{\chi}$ such that the associated coding X of Y is finite and non-ergodic (non-mixing).*

Proof. Consider the non-mixing case. By approximation with cylinder sets, there exist cylinder sets A, B such that, say, $\limsup_{t \rightarrow \infty} \mathbb{E}[\mathbf{1}_A(S^t Y) \mathbf{1}_B(Y)] > \mathbb{P}(A)\mathbb{P}(B)$, where $S : \mathcal{Y}^{\mathbb{Z}} \rightarrow \mathcal{Y}^{\mathbb{Z}}$ is the left shift. Define the time-zero coder $\tilde{\chi} : y \mapsto \mathbf{1}_{y \in A} + 2\mathbf{1}_{y \in B}$ (here, $\mathcal{X} = \{0, 1, 2, 3\}$), and let X be the corresponding coding of Y . Let $C \subset \mathcal{X}^{\mathbb{Z}}$ be the cylinder set defined as the preimage of $\{1, 3\}$ under the projection $x \mapsto x_0$, and D the preimage of $\{2, 3\}$. Then, if $T : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ is the left shift, $\mathbf{1}_C(T^t X) = \mathbf{1}_{\{1,3\}} \tilde{\chi}(S^t Y) = \mathbf{1}_A(S^t Y)$, and similarly $\mathbf{1}_D(X) = \mathbf{1}_B(Y)$. Hence, $\limsup_{t \rightarrow \infty} \mathbb{E}[\mathbf{1}_C(T^t X) \mathbf{1}_D(Y)] > \mathbb{P}(C)\mathbb{P}(D)$, and X cannot be mixing. The proof with non-ergodicity is similar. □

Convergence of the form (3) is preserved under finite codings of time series, which depend on only finitely many coordinates. Mixing and ergodicity are also preserved. Therefore, in light of the previous discussion, Lemma 1 has the following corollary, which states that (3) cannot generally hold with joinings.

Corollary 1. *Let $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ be stationary and non-mixing (non-ergodic), and let \mathfrak{Y} denote the set of random variables $Y : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ which are stationary and mixing (ergodic), and such that the joint distribution (X, Y) is a joining. Then there exist $t \in \mathbb{N}$ and $\varepsilon > 0$ such that $\sup_{Y \in \mathfrak{Y}} \mathbb{P}(X_{-t} = Y_{-t}, \dots, X_t = Y_t) < 1 - \varepsilon$.*

Although the convergence supplied by Theorems 1–3 does not generally imply convergence in the \bar{d} -metric, it does imply weak convergence of the law of X^k to X . Entropy is upper semicontinuous with respect to the weak topology when \mathcal{X} is finite [40, Theorem I.9.16], which is not generally true in non-compact settings (see [44, p. 184], [25], and references therein). However, upper semicontinuity of entropy does hold with greater generality when (3) is the mode of convergence.

The calculation of the joint entropy of a collection of random variables (X_1, \dots, X_t) takes on different forms depending on whether the variables are discrete or continuous, so we write a general form of joint entropy as

$$H(X_1, \dots, X_t) = - \int_{\mathcal{X}^t} f(x_1, \dots, x_t) \log f(x_1, \dots, x_t) \, d\mu^t, \tag{5}$$

where f is a density with respect to a Borel product measure μ^t on \mathcal{X}^t (e.g. the counting measure or Lebesgue measure [12]) and by convention $0 \log 0 = 0$. The entropy rate of a time series $X = (X_t)_{t \in \mathbb{Z}}$ is then defined by $H(X) = \lim_{t \rightarrow \infty} t^{-1} H(X_1, \dots, X_t)$. Provided that X_0 has finite entropy, we can show that entropy is upper semicontinuous with respect to the approximations provided by our results. Therefore, any collection of time series whose entropy is bounded away from 0 cannot approximate every time series in the sense of Theorems 1–3.

Lemma 2. *Let entropy H be given by (5). If \mathcal{X} is Polish and (X^k) and X are stationary \mathcal{X} -valued time series such that $X_0 \stackrel{d}{=} X_0^k$ for all k , $H(X_0) < \infty$, and (3) holds for all $t \in \mathbb{N}$, then $\limsup_{k \rightarrow \infty} H(X^k) \leq H(X)$.*

Proof. By stationarity, $H(X)$ is the decreasing limit of $H(X_0 | X_{-1}, \dots, X_{-t})$ [12], so just as in [40, Theorem I.9.1], it is sufficient to show that $H(X_0^k | X_{-1}^k, \dots, X_{-t}^k) \rightarrow H(X_0 | X_{-1}, \dots, X_{-t})$ for all t . To show convergence of the conditional entropies, it is enough to show convergence of the unconditional entropies $H(X_1^k, \dots, X_t^k)$ (cf. [12]). Fix a particular t .

Let $\Omega_0^k \subset \Omega$ be the set of ω such that $X_j^k(\omega) = X_j(\omega)$ for all j , $-t \leq j \leq t$. Let $\Omega_1^k = \Omega \setminus \Omega_0^k$. Suppose that $\mathbb{P}(\Omega_1^k) > 0$ (if not, there is nothing to prove). Let \mathbb{P}_t^k be the restriction of \mathbb{P} to Ω_t^k for $t \in \{0, 1\}$. Let f_t^k be the density on \mathcal{X}^t corresponding to the pushforward $(X_1^k, \dots, X_t^k)_* \mathbb{P}_t^k$, which exists by the Radon–Nikodym theorem, and let $-H_t(X_1^k, \dots, X_t^k) = \int_{\mathcal{X}^t} f_t^k \log f_t^k \, d\mu^t$ for $t \in \{0, 1\}$. Then, by the subadditivity of the map $x \mapsto -x \log x$,

$$\begin{aligned} H_0(X_1^k, \dots, X_t^k) - \mathbb{P}(\Omega_1^k) &\leq - \underbrace{\int_{\mathcal{X}^t} (f_0^k + f_1^k) \log (f_0^k + f_1^k) \, d\mu^t}_{=H(X_1^k, \dots, X_t^k)} \\ &\leq H_0(X_1^k, \dots, X_t^k) + H_1(X_1^k, \dots, X_t^k). \end{aligned}$$

A similar decomposition holds with $H(X_1, \dots, X_t)$, $H_0(X_1, \dots, X_t) = H_0(X_1^k, \dots, X_t^k)$, and $H_1(X_1, \dots, X_t)$. By assumption, $\mathbb{P}(\Omega_1^k) \rightarrow 0$, so it is sufficient to handle the H_1 terms. We show that any sequence of k contains a further subsequence k_n for which $\limsup_{n \rightarrow \infty} H_1(X_1^{k_n}, \dots, X_t^{k_n}) \leq 0$, which implies the desired convergence.

Let $\alpha_k = \mathbb{P}(\Omega_1^k)$, and let $f_1^k(x_j)$ denote the j th marginal of f_1^k . By considering the probability density f_1^k/α_k and letting $f(x)$ denote the density of X_0 , we have

$$\begin{aligned} H_1(X_1^k, \dots, X_t^k) &\leq - \sum_{j=1}^t \int_{\mathcal{X}} f_1^k(x_j) \log f_1^k(x_j) \, d\mu + (t-1)\alpha_k \log \alpha_k \\ &\leq - \sum_{j=1}^t \int_{\mathcal{X}} f_1^k(x_j) \log f_1^k(x_j) (\mathbf{1}_{f(x) \in [0, e^{-1}]} + \mathbf{1}_{f(x) \notin [0, e^{-1}]}) \, d\mu. \end{aligned} \tag{6}$$

The inequality $f_1^k(x_j) \leq f(x_j)$ holds, so $-f_1^k(x_j) \log f_1^k(x_j) \leq -f(x) \log f(x)$ when $f \in [0, e^{-1}]$. Also, $\int_{\mathcal{X}} \mathbf{1}_{f(x) \notin [0, e^{-1}]} \, d\mu \leq e$. Now, by passing to a subsequence if necessary, we can assume that $f_1^k(x_j)$ converges pointwise to 0 by the fact that $\alpha_k \rightarrow 0$. Thus, the dominated convergence theorem implies that both pieces of (6) converge to 0. A similar argument applies to $H_1(X_1, \dots, X_t)$, which concludes the proof. \square

2.2. Testability

Let us formalize a notion of statistical testing in our setting. Let X be an \mathcal{X} -valued stationary time series with law $\mathbb{P} \in \mathbf{P}$ (where the set \mathbf{P} is to be specified later) on the space $\mathcal{X}^{\mathbb{Z}}$. We are interested in hypothesis testing problems of the form [8]

$$H_0 : \mathbb{P} \in \mathbf{P}_0, \quad H_1 : \mathbb{P} \in \mathbf{P}_1,$$

where $\mathbf{P}_0 \subset \mathbf{P}$ is the set of probability measures on Ω for which the null hypothesis is deemed to hold, and $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$ is its complement. A possibly randomized statistical test will be denoted by a map $\varphi_t : \mathcal{X}^t \rightarrow [0, 1]$, where t indicates the number of time periods observed. The corresponding size of the test is given by

$$\sup_{\mathbb{P} \in \mathbf{P}_0} \mathbb{E}[\varphi_t(X_1, \dots, X_t)] = \sup_{\mathbb{P} \in \mathbf{P}_0} \int_{\Omega} \varphi_t(x_1, \dots, x_t) \, dP(x).$$

We will show that when $\mathbf{P}_0, \mathbf{P}_1$ are chosen to test the ergodicity (or non-ergodicity) of the time series X , we have, under mild conditions,

$$\sup_{\mathbb{P} \in \mathbf{P}_1} \mathbb{E}_{\mathbb{P}}[\varphi_t] \leq \sup_{\mathbb{P} \in \mathbf{P}_0} \mathbb{E}_{\mathbb{P}}[\varphi_t] \tag{7}$$

for any test φ_t and any sample size t . In other words, the power of the test φ_t cannot exceed its size, for any t . For this reason, we can say that the null hypothesis H_0 is non-testable.

The definition of \mathbf{P} and its constituent sets \mathbf{P}_0 and \mathbf{P}_1 clearly play a significant role in establishing (7). We let \mathbf{P} denote the set of stationary Borel measures on $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty})$. We then let $\mathbf{P}_E \subset \mathbf{P}$ denote the subset of distributions which correspond to \mathcal{X} -valued time series which are additionally ergodic, and \mathbf{P}_M those which are mixing. In the special case where $\mathcal{X} = \mathbb{R}$ and \mathbf{P} contains only distributions induced by time series X which are integrable in the sense that $\mathbb{E}[|X_0|] < \infty$, we let $\mathbf{P}_{ME} \subset \mathbf{P}$ denote the subset of distributions which correspond to mean ergodic time series. By Corollary 2, $\mathbf{P}_E \subset \mathbf{P}_{ME}$.

We consider several null hypotheses. The first sets $\mathbf{P}_0 = \mathbf{P}_E$. This is to say that the time series in question is ergodic, under the null hypothesis. The other scenarios we consider are that $\mathbf{P}_0 = \mathbf{P}_M, \mathbf{P}_E^c, \mathbf{P}_{ME},$ and \mathbf{P}_{ME}^c . We show that (7) holds in all scenarios.

The specific results that we use in the proof of Theorem 1 depend on the aperiodicity of the underlying transformation, which by analogy might suggest that our results are valid only for aperiodic time series. However, Theorem 1 shows not only that periodicity is not a hindrance to ergodic approximation, but that the approximation can always be done with aperiodic series. This has the added benefit of implying that all of our non-testability results still apply when we restrict attention to only the set of aperiodic time series. As aperiodicity is a light assumption, this negates what would otherwise be a straightforward workaround to our finding of non-testability.

As a corollary to Theorem 1, we have our main testability result.

Corollary 2. *Let \mathcal{X} be a Polish space satisfying $|\mathcal{X}| \geq 2$. Then, for any $t \in \mathbb{N}$, (7) holds with $\mathbf{P}_0 = \mathbf{P}_E$, \mathbf{P}_M , and \mathbf{P}_E^c for all tests φ_t . If $\mathcal{X} \subset \mathbb{R}$ and \mathbf{P} contains only integrable distributions (corresponding to time series X satisfying $\mathbb{E}[|X_0|] < \infty$), then (7) holds with $\mathbf{P}_0 = \mathbf{P}_{ME}$ and $\mathbf{P}_0 = \mathbf{P}_{ME}^c$.*

The same is true if \mathbf{P} , \mathbf{P}_0 , and \mathbf{P}_1 are additionally restricted to the set of aperiodic distributions.

Proof. We deal with the case $\mathbf{P}_0 = \mathbf{P}_E$; the other cases are similar. Let φ_t be a test function. By Theorem 1 there exists a sequence (X^k) of stationary and ergodic processes such that $\lim_{k \rightarrow \infty} \mathbb{P}(X_1 = X_1^k, \dots, X_t = X_t^k) = 1$. Let \mathbb{P}^k denote the specific probability law of (X^k) , and \mathbb{P} the law of X . Then, by the boundedness of φ_t ,

$$\sup_{\mathbb{P} \in \mathbf{P}_E} \mathbb{E}_{\mathbb{P}}[\varphi_t] \geq \lim_{k \rightarrow \infty} \mathbb{E}_{\mathbb{P}^k}[\varphi_t] = \lim_{k \rightarrow \infty} \mathbb{E}[\varphi_t(X_1^k, \dots, X_t^k)] = \mathbb{E}[\varphi_t(X_1, \dots, X_t)] = \mathbb{E}_{\mathbb{P}}[\varphi_t].$$

As φ_t and X , whence \mathbb{P} , were arbitrary, (7) is proved. \square

The union bound can be applied to show that the non-testability result holds even if φ_t is allowed to depend not only on X_1, \dots, X_t , but an independent and identically distributed (i.i.d.) sample $(X_1^i, \dots, X_t^i)_{i=1}^n$ of size n , where n is arbitrarily large. The approximation in Theorem 1 is quite strong, so the non-testability result can be extended via the union bound to other situations in which the sample is not necessarily i.i.d. This contrasts with, say, [13, 14], which consider Markovian time series and construct tests for stationary ergodicity that depend upon the user's ability to 'draw' several i.i.d. observations of the series.

If aperiodicity of the approximating time series is dropped, Theorem 2 shows that any property of an aperiodic time series that is preserved under coding (e.g. ergodicity, mixing) of a time series must be dense in the sense of (3). Therefore, the null hypothesis that such a property holds is not testable. By contrapositive, any property for which a consistent statistical test is available must not be preserved under coding of a time series. Theorem 3 implies a similar failure for the testability of periodicity.

3. The category of stationary series

We now parallel, and invoke, results on the approximation and genericity of m.p. transformations possessing certain mixing properties (especially [16, 22]) to show the genericity of weak mixing in a large space of time series. We also find that the set of mixing processes is meagre, which augments analogous results [17, 35] that the set of mixing and stationary measures (or transformations, cf. [37]) is meagre in the weak topology. This supports the plausibility of the ergodic assumption and weak mixing for non-parametric time series but suggests that more restrictive settings are better suited to invoking mixing or even stronger notions.

In brief, we show that classical Baire category characterizations of mixing measures [17, 35] and transformations [15, 22, 37] carry over to a large metric space \mathbb{X} of time series:

$$\underbrace{\text{i.i.d.} \implies \varphi, \beta, \rho, \alpha \text{ mixing} \implies \text{mixing}}_{\text{meagre: Proposition 2}} \implies \underbrace{\text{weak mixing} \implies \text{ergodic} \implies \text{mean ergodic}}_{\text{generic: Proposition 1}}. \tag{8}$$

It will be necessary to define a structure for the time series under consideration so that we may properly define, e.g., their joint distributions. The space of processes that we define is quite large, consisting of maps from countable products of standard probability spaces into a sequence space. The space is closed under extensions of the underlying probability space, as in (2). If this accommodation is not made, then even density results in the sense of Theorem 1 are not possible; Remark 1 implies that the aperiodic processes, viewed as a subset of time series mapping $\Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$, fail to be dense so long as there exists an injection from Ω to \mathcal{X} .

As before, let \mathcal{X} be a Polish space with its Borel σ -algebra and, for all $z \in \mathbb{R}$, let $(\Omega_z, \mathfrak{F}_z, \mathbb{P}_z)$ be a standard probability space. We consider stationary time series X that map either a finite or countably infinite collection, indexed by $V(X)$, of these probability spaces into $\mathcal{X}^{\mathbb{Z}}$. Accordingly, let $\tilde{\mathbb{X}}$ denote the set of stationary time series $X : (\prod_{z \in V(X)} \Omega_z, \prod_{z \in V(X)} \mathfrak{F}_z, \prod_{z \in V(X)} \mathbb{P}_z) \rightarrow \mathcal{X}^{\mathbb{Z}}$, where $V(X) \subset \mathbb{R}$ indexes the domain $\prod_{z \in V(X)} \Omega_z$ of X and is either finite or countable. The requirement $V(X) \subsetneq \mathbb{R}$ is to ensure that, given $X \in \tilde{\mathbb{X}}$, there always exists a time series $Y \in \tilde{\mathbb{X}}$ independent of X .

With the convention established in (2), we can define any two elements of $\tilde{\mathbb{X}}$ to lie on a common probability space as follows. Given elements X and Y of $\tilde{\mathbb{X}}$ and any $\ell \in \mathbb{N}$, define $\pi_{X,Y}^X : \prod_{z \in V(X) \cup V(Y)} \Omega_z \rightarrow \prod_{z \in V(X)} \Omega_z$ to be the projection onto the $\Omega_z, z \in V(X)$, which appear in the domain $\prod_{z \in V(X)} \Omega_z$ of X , and define $\pi_{X,Y}^Y$ similarly. We may define a pseudometric d_ℓ on the space $\tilde{\mathbb{X}}$ by

$$d_\ell(X, Y) = \mathbb{P}\left((X_{-\ell}, \dots, X_\ell) \circ \pi_{X,Y}^X \neq (Y_{-\ell}, \dots, Y_\ell) \circ \pi_{X,Y}^Y \right),$$

where \mathbb{P} is used to denote the product measure $\prod_{z \in V(X) \cup V(Y)} \mathbb{P}_z$ on $\prod_{z \in V(X) \cup V(Y)} \Omega_z$. These pseudometrics may be combined to form a pseudometric d on the space $\tilde{\mathbb{X}}$ that metrizes (3): $d(X, Y) = \sum_{\ell=1}^\infty 2^{-\ell} d_\ell(X, Y)$. Finally, let \sim denote the equivalence relation induced by the d pseudometric, and let $\mathbb{X} = \tilde{\mathbb{X}}/\sim$ denote the resulting quotient space. This is the quotient space obtained by imposing the relation indicated in (2) on $\tilde{\mathbb{X}}$. We refer to equivalence classes $[X] \in \mathbb{X}$ by writing simply X . The resulting space is Baire.

Lemma 3. (\mathbb{X}, d) is a complete metric space.

Proof. The only difficulty in verifying that d is a pseudometric on $\tilde{\mathbb{X}}$ lies in checking the triangle inequality. Pick $(X, V), (Y, V'), (Z, V'') \in \tilde{\mathbb{X}}$ arbitrarily and note that

$$\begin{aligned} d_\ell((X, V), (Y, V')) &= \mathbb{P}\left((X_{-\ell}, \dots, X_\ell) \circ \pi_{V \cup V' \cup V''}^V \neq (Y_{-\ell}, \dots, Y_\ell) \circ \pi_{V \cup V' \cup V''}^{V'} \right) \\ &\leq \mathbb{P}\left((X_{-\ell}, \dots, X_\ell) \circ \pi_{V \cup V' \cup V''}^V \neq (Z_{-\ell}, \dots, Z_\ell) \circ \pi_{V \cup V' \cup V''}^{V''} \right) \\ &\quad + \mathbb{P}\left((Y_{-\ell}, \dots, Y_\ell) \circ \pi_{V \cup V' \cup V''}^{V'} \neq (Z_{-\ell}, \dots, Z_\ell) \circ \pi_{V \cup V' \cup V''}^{V''} \right) \\ &= d_\ell((X, V), (Z, V'')) + d_\ell((Y, V'), (Z, V'')). \end{aligned}$$

It remains to show completeness. Let (X^k, V^k) be a d -Cauchy sequence of equivalence classes in \mathbb{X} . Choose a subsequence $(k_n)_{n \in \mathbb{N}}$ such that, for all ℓ ,

$$\sum_{n=1}^{\infty} d_{\ell}((X^{k_{n+1}}, V^{k_{n+1}})(X^{k_n}, V^{k_n})) < \infty.$$

Then the Borel–Cantelli lemma implies that, for each $\ell \in \mathbb{N}$,

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \left\{ (X_{-\ell}^{k_{n+1}}, \dots, X_{\ell}^{k_{n+1}}) \circ \pi_{\bigcup_{i \in \mathbb{N}} V_i}^{V^{k_{n+1}}} \neq (X_{-\ell}^{k_n}, \dots, X_{\ell}^{k_n}) \circ \pi_{\bigcup_{i \in \mathbb{N}} V_i}^{V^{k_n}} \right\}\right) = 0,$$

so that, almost surely, $\lim_{n \rightarrow \infty} (X_{-\ell}^{k_n}, \dots, X_{\ell}^{k_n}) \circ \pi_{\bigcup_{i \in \mathbb{N}} V_i}^{V^{k_n}}$ exists as a random variable defined on $\prod_{z \in \bigcup_{i \in \mathbb{N}} V_i} \Omega_z$. Define $X_t = \lim_{n \rightarrow \infty} X_t^{k_n} \circ \pi_{\bigcup_{i \in \mathbb{N}} V_i}^{V^{k_n}}$ for all $t \in \mathbb{Z}$ and let X denote the resulting time series defined on $V = \bigcup_{i \in \mathbb{N}} V_i$, so that in \mathbb{X} the convergence $(X^k, V^k) \rightarrow (X, V)$ holds. Because the convergence $(X^{k_n}) \rightarrow X$ holds in the product topology of $\mathcal{X}^{\mathbb{Z}}$, and $\mathcal{X}^{\mathbb{Z}}$ is Polish with this topology, X is measurable. Stationarity of X follows from stationarity of (X_t^k) . \square

Now let \mathbb{X}_{WM} denote the subset of \mathbb{X} consisting of only weakly mixing transformations. Then, \mathbb{X}_{WM} is G_{δ} . The following lemma adapts the strategy of [22] to time series.

Lemma 4. \mathbb{X}_{WM} is a G_{δ} subset of (\mathbb{X}, d) .

Proof. For each $\ell \in \mathbb{N}$ let $C_{\ell} = \mathcal{X}^{2\ell+1}$ with its product Borel σ -algebra, and let $\pi_{\ell} : \mathcal{X}^{\mathbb{Z}} \rightarrow C_{\ell}$ denote the projection $(\dots, x_{-\ell}, \dots, x_{\ell}, \dots) \mapsto (x_{-\ell}, \dots, x_{\ell})$. C_{ℓ} is a separable metric space, so it is second countable with a basis which we can denote as $\{B_i^{\ell}\}_{i \in \mathbb{N}}$. The collection $\{\pi_{\ell}^{-1}(B_i^{\ell})\}_{i, \ell \in \mathbb{N}}$ then constitutes a basis for the product topology on $\mathcal{X}^{\mathbb{Z}}$. Let \mathcal{A} denote the algebra which is generated by the family of sets $\{\pi_{\ell}^{-1}(B_i^{\ell})\}_{i, \ell \in \mathbb{N}}$, which is again countable (as the family of sets obtained from the countable basis sets by iterating complementation and finite unions finitely many times). The σ -algebra generated by \mathcal{A} contains all of the open sets of the product topology in $\mathcal{X}^{\mathbb{Z}}$, so it is the Borel σ -algebra \mathfrak{B} on $\mathcal{X}^{\mathbb{Z}}$. Furthermore, by the density of algebras in the σ -algebras they generate, for any Borel set $D \subset \mathcal{X}^{\mathbb{Z}}$, constant $\varepsilon > 0$, and finite Borel measure μ on $\mathcal{X}^{\mathbb{Z}}$, there exists a set $B \in \mathcal{A}$ satisfying $\mu(A \Delta B) = \mu(A \setminus B) + \mu(B \setminus A) < \varepsilon$. Hence, letting Λ denote the countable set of finite linear combinations of functions $\{\mathbf{1}_B\}_{B \in \mathcal{A}}$ with coefficients in $\mathbb{Q} + i\mathbb{Q}$, Λ is dense in $L^2(\mathcal{X}^{\mathbb{Z}}, \mu)$. Let f_1, f_2, \dots be an enumeration of Λ .

Now, using the notation of [22], for all i, j, m, t , define

$$E(i, j, m, n) = \{X \in \mathbb{X} : |\mathbb{E}[f_i(S^n(X))\overline{f_j(X)}] - \mathbb{E}[f_i(X)]\mathbb{E}[\overline{f_j(X)}]| < 1/m\},$$

where $S : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ is the left shift map.

First, we argue that $E(i, j, m, n)$ is open in (\mathbb{X}, d) . Define the function $\gamma_{i,j,n} : \mathbb{X} \rightarrow \mathbb{C}$ by

$$\gamma_{i,j,n} : X \mapsto \mathbb{E}[f_i(S^n(X))\overline{f_j(X)}] - \mathbb{E}[f_i(X)]\mathbb{E}[\overline{f_j(X)}].$$

By the construction of Λ , the function f_i only depends on the finite-dimensional distribution of $\pi_{\ell} X = (X_{-\ell}, \dots, X_{\ell})$ for some ℓ sufficiently large. Let X^k converge to X in \mathbb{X} ; then the distribution of $\pi_{\ell}(X^k)$ converges to that of $\pi_{\ell}(X)$ in the total variation norm. In particular, $\mathbb{E}[f_i(X^k)] \rightarrow \mathbb{E}[f_i(X)]$, so $\mathbb{E}[f_i(\cdot)] : \mathbb{X} \rightarrow \mathbb{C}$ is d -continuous. Similarly, $\gamma_{i,j,n}$ is d -continuous, so that $E(i, j, m, n) = \gamma_{i,j,n}^{-1}\{z \in \mathbb{C} : |z| < 1/m\}$ is open.

We claim that $\mathbb{X}_{\text{WM}} = \bigcap_{i,j,m \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} E(i, j, m, n)$. By expressing f_i and f_j as sums over indicator functions of cylinder sets of the form $\pi_\ell^{-1}(B_i^\ell)$, it is clear that (1) implies containment in the forward direction. Conversely, suppose that X is not weakly mixing, so that (1) does not hold. Let $\tilde{\mathbb{P}}$ denote the pushforward measure induced on $\mathcal{X}^{\mathbb{Z}}$ by X so that, by applying [41, Proposition 6.2.2] with the roles of A and B reversed, in conjunction with the stationarity of X , the left shift S is not weakly mixing on $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}, \tilde{\mathbb{P}})$. Because Λ is dense in $L^2(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$, it follows as in [22] that there exists an $f_i \in \Lambda$ such that

$$\gamma_{i,i,n}(X) = \left| \int_{\mathcal{X}^{\mathbb{Z}}} f_i(S^n \tilde{\omega}) \overline{f_i(\tilde{\omega})} \, d\tilde{\mathbb{P}} - \int f_i(\tilde{\omega}) \, d\tilde{\mathbb{P}} \int \overline{f_i(\tilde{\omega})} \, d\tilde{\mathbb{P}} \right| > \frac{1}{2}$$

for all n , which concludes the proof. □

As Theorem 1 implies the density of \mathbb{X}_{WM} in \mathbb{X} , we immediately have the following result.

Proposition 1. \mathbb{X}_{WM} is a residual subset of (\mathbb{X}, d) .

Now let \mathbb{X}_{M} denote the set of mixing processes in \mathbb{X} . Let \mathfrak{M} denote the set of stationary measures on $\mathcal{X}^{\mathbb{Z}}$ equipped with the weak topology, and let $\mathfrak{M}_{\text{M}} \subset \mathfrak{M}$ be the subset of measures which are mixing. \mathfrak{M} is a separable metric space under this topology, and [35, Theorem 3.4] establishes that \mathfrak{M}_{M} is of the first Baire category in \mathfrak{M} . Note that the mapping $\psi : \mathbb{X} \rightarrow \mathfrak{M}$ given by $\psi : (X, V) \mapsto X_* \prod_{z \in V} \mathbb{P}_z$ is continuous with respect to the d -metric and the weak topology on \mathfrak{M} , and that $\mathbb{X}_{\text{M}} = \psi^{-1}(\mathfrak{M}_{\text{M}})$, $\mathbb{X}_{\text{M}}^c = \psi^{-1}(\mathfrak{M}_{\text{M}}^c)$ (by definition). It follows that \mathbb{X}_{M}^c is a G_δ subset of \mathbb{X} , and Theorem 1 implies that it is dense (as are, in fact, the set of non-ergodic processes), giving the following result.

Proposition 2. \mathbb{X}_{M} is meagre in (\mathbb{X}, d) .

Remark 2. In the same way, Theorem 1 can be applied to establish (8) for stronger topologies on the space $\mathcal{M}(\mathcal{X}^{\mathbb{Z}})$ of stationary measures on $\mathcal{X}^{\mathbb{Z}}$ (existing work has largely focused on the weak topology). For instance, consider the weakest topology \mathcal{T} on $\mathcal{M}(\mathcal{X}^{\mathbb{Z}})$ that makes the map $\mu \mapsto \int f \, d\mu$ continuous for every bounded f . Let $f : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathbb{R}$ be bounded, and let $X^k \rightarrow X$ in \mathbb{X} . Let μ^k be the law of X^k and μ the law of X . Inasmuch as f can be approximated above and below with simple functions constructed over cylinder sets, we have $\int_{\mathcal{X}^{\mathbb{Z}}} f \, d\mu^k = \mathbb{E}[f(X^k)] \rightarrow \mathbb{E}[f(X)] = \int_{\mathcal{X}^{\mathbb{Z}}} f \, d\mu$. Therefore, the map which sends points in \mathbb{X} to their distributions in $(\mathcal{M}(\mathcal{X}^{\mathbb{Z}}), \mathcal{T})$ is a continuous surjection. We conclude that the image of any dense subset of \mathbb{X} under this map is itself dense. In conjunction with existing genericity results [17, 35], this implies that (8) holds for $(\mathcal{M}(\mathcal{X}^{\mathbb{Z}}), \mathcal{T})$.

4. Proofs of main results

Here, we give proofs for Theorems 1–3. Additional proofs are gathered in the supplementary appendix.

Proof of Theorem 1. Let \mathfrak{B} be the Borel σ -algebra on a Polish space \mathcal{X} . Define the product space $\mathcal{X}^{\mathbb{Z}} \equiv \prod_{t \in \mathbb{Z}} \mathcal{X}$ with the corresponding product Borel σ -algebra $\bigotimes_{t \in \mathbb{Z}} \mathfrak{B}$. Note that this σ -algebra, which we denote \mathfrak{B}^∞ , is also the Borel σ -algebra on $\mathcal{X}^{\mathbb{Z}}$ [27, Lemma 1.2] equipped with its product topology. Let $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ be Borel measurable, where the projection of X onto its t th component is denoted $X_t : \Omega \rightarrow \mathcal{X}$ for all $t \in \mathbb{Z}$. Then we have the following result.

Lemma 5. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ be Borel measurable, and let $\tilde{\mathbb{P}}$ denote the pushforward measure $X_*\mathbb{P}$. Then $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ is a standard probability space, and if $\mathbb{P}(X = x) = 0$ for every element $x \in \mathcal{X}^{\mathbb{Z}}$, $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ is isomorphic to Lebesgue measure on the unit interval $[0, 1]$.*

Proof. Note that $\mathcal{X}^{\mathbb{Z}}$ is a countable product of Polish spaces with the product topology, and thus is itself Polish [42, §2.2]. By [7, Theorem 7.1.7], $\tilde{\mathbb{P}}$ is regular. Hence, [26, Theorem 2.4.1] implies that $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ is also a standard probability space, using again the fact that $\mathcal{X}^{\mathbb{Z}}$ is Polish with its product topology.

As $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ is standard, [7, Theorem 9.4.7] implies that it is isomorphic to the unit interval equipped with its Borel σ -algebra and Lebesgue measure, and a countable number of atoms (call this space Y). Let $f : \mathcal{X}^{\mathbb{Z}} \rightarrow Y$ be an isomorphism of these measure spaces. If the pushforward $f_*\tilde{\mathbb{P}}$ has an atom $A \in Y$, then $X(f^{-1}(A))$ is an atom for the Borel measure $\tilde{\mathbb{P}}$ on $\mathcal{X}^{\mathbb{Z}}$, and must be a singleton. So if \mathbb{P} assigns no singletons positive measure then $\mathcal{X}^{\mathbb{Z}}$ is isomorphic to Lebesgue measure on the unit interval. \square

We now discuss some measure-theoretical preliminaries following [5, 41]. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space. Whenever two sets A and B are equal up to a set of measure zero, i.e. $\mathbb{P}(A \Delta B) = 0$, we write $A = B \pmod{\mathbb{P}}$. A *transformation* is a measurable map $T : (\Omega, \mathfrak{F}, \mathbb{P}) \rightarrow (\Omega, \mathfrak{F}, \mathbb{P})$. In this paper, we only consider transformations with measurable inverses. T is *non-singular* if $\mathbb{P}(A) = 0$ if and only if $\mathbb{P}(T^{-1}(A)) = 0$. We say that T is *measure-preserving* (m.p.) if $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ for all measurable A . T is *ergodic* if, for all A with $\mathbb{P}(A) > 0$ such that $T^{-1}(A) = A \pmod{\mathbb{P}}$ we have $\mu(A) = 1$ or $\mu(A^c) = 0$ [41, Lemma 3.7.1]. An m.p. transformation is *weakly mixing* if, for each pair of measurable sets A and B , there exists some zero-density set $I_0 \subset \mathbb{N}$ such that

$$\lim_{t \rightarrow \infty, t \notin I_0} \mathbb{P}(T^{-j}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{9}$$

[41, Proposition 6.2.2] provides a number of equivalent characterizations of weak mixing. T is *mixing* if (9) holds with $I_0 = \emptyset$. Evidently, mixing implies weak mixing, which in turn implies ergodicity.

Let $(\Omega, \mathfrak{F}, \mathbb{P})$ denote a non-atomic measure space, and let $G(\Omega, \mathbb{P})$ denote the group of all non-singular transformations on it; furthermore, let $M(\Omega, \mathbb{P}) \subset G(\Omega, \mathbb{P})$ denote the subgroup of m.p. transformations on Ω (see the introduction of [2]). The uniform topology on $G(\Omega, \mathbb{P})$ is induced by the metric $d(T, S) = \mathbb{P}(\omega : T(\omega) \neq S(\omega))$. This topology is complete and includes $M(\Omega, \mathbb{P})$ as a closed subset of $G(\Omega, \mathbb{P})$, so $G(\Omega, \mathbb{P})$ and $M(\Omega, \mathbb{P})$ are complete metric spaces and hence Baire spaces.

Say that a point $\omega \in \Omega$ is a *periodic point* of period t for T if $T^t x = x$ and $T^j x \neq x$ for $j = 1, \dots, t - 1$. A transformation T is called *aperiodic* if the set of its periodic points has \mathbb{P} -measure 0. We use $\mathcal{A}p$ to denote the set of aperiodic transformations in $M(\Omega, \mathbb{P})$.

Using Lemma 5, we have the following equivalence, which is given in the case $\mathcal{X} = \mathbb{R}$ in [28] and can be proved with the π - λ theorem [6].

Proposition 3. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and $X : \Omega \rightarrow \mathcal{X}^{\mathbb{Z}}$ be Borel measurable and non-atomic in the sense of Lemma 5. Then $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ (where $\tilde{\mathbb{P}}$ is the pushforward $X_*\mathbb{P}$) is a standard non-atomic probability space, and X is stationary if and only if the left shift map $T : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ is an m.p. transformation thereon. In this case,*

$$X_t(\omega) = \chi(T^t \circ X(\omega)), \tag{10}$$

where $\chi : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}$ is the projection onto the zeroth coordinate. Moreover, under stationarity of X , this T satisfies:

- (i) The series is mixing (ergodic) if and only if T is a mixing (ergodic) transformation.
- (ii) The series is aperiodic if and only if $T \in \text{Ap}$.

Proposition 3 applies to the left shift map on $\mathcal{X}^{\mathbb{Z}}$. It is not generally true that the time series $\chi(T^t \circ \omega)$ is non-mixing (non-ergodic) if T is non-mixing (non-ergodic), although the converse is true. As $\tilde{\mathbb{P}} = X_*\mathbb{P}$, we have the following corollary.

Corollary 3. *In the setting of Proposition 3, a stationary time series $(X_t(\omega))_{t \in \mathbb{Z}}$ is ergodic if and only if, for all $g \in L^1(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n g((\dots, X_{-1+j}, X_j, X_{1+j}, \dots)) \stackrel{\text{a.s.}}{=} \mathbb{E}_{\tilde{\mathbb{P}}}[g((\dots, X_{-1}(\omega), X_0(\omega), X_1(\omega), \dots))] = \mathbb{E}_{\tilde{\mathbb{P}}}[g]. \tag{11}$$

The next lemma demonstrates that convergence in the uniform topology implies convergence of a joint distribution.

Lemma 6. *Let $T_k \rightarrow T$ in $G(\Omega, \mathbb{P})$ equipped with the uniform topology. Then*

$$\mathbb{P}((T_k^{-t}\omega, \dots, T_k^t\omega) = (T^{-t}\omega, \dots, T^t\omega)) \rightarrow 1. \tag{12}$$

Proof. We claim that $\bigcap_{j=0}^{t-1} T^j(A_k) \subset \{\omega : T_k^j\omega = T^j\omega \text{ for all } j = 1, \dots, t\}$, where $A_k = \{\omega : T_k(\omega) = T\omega\}$. Indeed, if $\omega \in \bigcap_{j=0}^{t-1} T^{-j}(A_k)$ then we have $T^j\omega \in A_k$ for all $j = 0, \dots, t-1$, which is to say that $T_k(\omega) = T\omega, T_k(T\omega) = T^2\omega, \dots, T_k(T^{t-1}\omega) = T^t\omega$, which, by substitution, implies that $T_k^j(\omega) = T^j\omega$ for $j = 1, \dots, t$. Now note that $\mathbb{P}(T^j A_k) = \mathbb{P}(A_k) \rightarrow 1$ for all j , so that $\mathbb{P}((T_k\omega, \dots, T_k^t\omega) = (T\omega, \dots, T^t\omega)) \rightarrow 1$. To finish, note that, because T and T_k are invertible and m.p.,

$$\mathbb{P}(T\omega = T_k\omega) = \mathbb{P}(\omega = T^{-1}T_k\omega) = \mathbb{P}(T_k\omega : \omega = T^{-1}T_k\omega) = \mathbb{P}(\omega' : T_k^{-1}\omega' = T^{-1}\omega'),$$

so that also $T_k^{-1} \rightarrow T^{-1}$ uniformly; now apply the same argument for $j = -1, \dots, -t$ and apply the union bound. □

It is also useful to have the following lemma in hand, which assures us that we may replace the assumption that X follows a non-atomic distribution with the slightly stronger assumption that it is aperiodic.

Lemma 7. *An aperiodic \mathcal{X} -valued time series X is also non-atomic in the sense of Lemma 5.*

Proof. We prove the contrapositive. Let \mathbb{P} be the measure induced by X on $\mathcal{X}^{\mathbb{Z}}$. Suppose that \mathbb{P} has an atom $x \in \mathcal{X}^{\mathbb{Z}}$ so that $\mathbb{P}(x) = \mathbb{P}(X_t = x_t \text{ for all } t \in \mathbb{Z}) > 0$. As X_t is stationary, for all $s \in \mathbb{Z}$ we have $\mathbb{P}(T^s x) = \mathbb{P}(X_t = x_{t+s} \text{ for all } t \in \mathbb{Z}) = \mathbb{P}(X_t = x_t \text{ for all } t \in \mathbb{Z}) > 0$. If $T^s x \neq T^{s'} x$ whenever $s \neq s'$ then this contradicts that \mathbb{P} is a probability (finite) measure, so that for some $s \neq s'$ we have $T^{s-s'} x = x$. Assuming without loss of generality that $s > s'$, x is periodic with period $s - s'$ and X is not aperiodic. □

Proposition 4. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and \mathcal{X} a Polish space. Then, for any $t \in \mathbb{N}$ and stationary, aperiodic \mathcal{X} -valued time series X on $(\Omega, \mathfrak{F}, \mathbb{P})$, there is a sequence (X^k) of*

stationary, aperiodic, and mixing \mathcal{X} -valued time series supported on $(\Omega, \mathfrak{F}, \mathbb{P})$ such that $X_0^k = X_0$ for all $k \in \mathbb{N}$ and

$$\lim_{k \rightarrow \infty} \mathbb{P}(X_{-t} = X_{-t}^k, \dots, X_t = X_t^k) = 1.$$

The same is true if, instead of mixing, (X^k) is specified to be non-ergodic. If X is, in addition, an integrable \mathbb{R} -valued time series, the same is true if instead of mixing, (X^k) is specified to be mean-ergodic or non-mean-ergodic.

Proof. Fix $t \in \mathbb{N}$. We break the proof into four cases.

Case 1: Approximation with mixing processes. Fix the law \mathbb{P} corresponding to the \mathcal{X} -valued time series X which is stationary and aperiodic, as well as non-atomic by Lemma 7. Proposition 3 holds, and so $X_t = \chi(T^t \circ X(\omega))$ with T the left shift map and χ the projection onto the 0th coordinate. For brevity, we will typically write $\tilde{\omega}$ in place of $X(\omega)$. Let $\tilde{\mathbb{P}}$ denote the pushforward $X_*\mathbb{P}$. Lemmas 7 and 5 imply that $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^\infty, \tilde{\mathbb{P}})$ is a standard non-atomic measure space, and Proposition 3 guarantees that $T \in \mathcal{A}p \subset M(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$.

In [16, Theorem 7.14], it is shown that the conjugacy class of any mixing transformation $S \in \mathcal{A}p \cap M(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$ is dense in $\mathcal{A}p$, which is to say that there is a sequence of $\tilde{\mathbb{P}}$ -m.p. transformations $(\eta_k)_{k \in \mathbb{N}}$ such that $T_k \equiv \eta_k^{-1} S \eta_k \rightarrow T$ (uniform topology). Mixing is an isomorphic invariant [44, Theorem 2.13], so T_k is mixing.

Consider for each k the time series X^k defined by $X_t^k \equiv \chi(T_k^t \circ X(\omega))$ for $t \in \mathbb{Z}$. We immediately have $X_0^k = X_0$ for all k . Because $T_k : \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ is a Borel map, X^k measurably maps Ω to \mathcal{X} . Moreover, X^k is a stationary time series by Proposition 3, and the process is mixing by application of (9) to the definition of mixing given in (1) (with $I_0 = \emptyset$).

We now wish to show that X^k is aperiodic. Suppose for the sake of contradiction that the time series is not aperiodic, so that on some positive $\tilde{\mathbb{P}}$ -measure set $A \subset \mathcal{X}^{\mathbb{Z}}$, $\tilde{\omega} \in A$ implies that $\chi(T_k^t \tilde{\omega})$ is periodic. By partitioning A into a countable number of subsets, there must exist a positive measure subset $P_m \subset A$ on which X^k is periodic with period $m \in \mathbb{N}$, i.e. $\tilde{\omega} \in P_m \implies \chi(T_k^t \tilde{\omega}) = \chi(T_k^{t+m} \tilde{\omega})$ for all $t \in \mathbb{Z}$.

Note that $X_0^k(\tilde{\omega}) = \chi(\tilde{\omega})$ has the same distribution as X_0 . We claim that this implies the existence of two disjoint sets $B_1, B_2 \subset \mathcal{X}$ such that $\tilde{\mathbb{P}}(\chi^{-1}(B_1) \cap P_m) > 0$ and $\tilde{\mathbb{P}}(\chi^{-1}(B_2)) > 0$. In fact, this follows from the separability of \mathcal{X} ; for all $\varepsilon > 0$ there is a pairwise-disjoint cover $(C_\ell^\varepsilon)_{\ell \in \mathbb{N}}$ of \mathcal{X} such that (C_ℓ^ε) is contained in an ε -ball of \mathcal{X} and $\mathcal{X} = \bigsqcup_{\ell \in \mathbb{N}} C_\ell^\varepsilon$. Letting ε tend towards 0, there must eventually exist two sets $C_\ell^\varepsilon, C_{\ell'}^\varepsilon$ satisfying $\tilde{\mathbb{P}}(\chi^{-1}(C_\ell^\varepsilon)), \tilde{\mathbb{P}}(\chi^{-1}(C_{\ell'}^\varepsilon)) > 0$, or else because \mathcal{X} is Polish, the pushforward measure $\chi_*\tilde{\mathbb{P}}$ is a point mass at some $x_0 \in \mathcal{X}$. This would then imply that $\mathbb{P}(X_0 = x_0) = \mathbb{P}(\omega : \chi(\omega) = x_0) = 1$, and then by the stationarity of X that $\mathbb{P}(X_t = x_0 \text{ for all } t \in \mathbb{Z}) = 1$, contradicting that $\tilde{\mathbb{P}}$ is non-atomic. As we may choose C_ℓ^ε satisfying $\tilde{\mathbb{P}}(\chi^{-1}(C_\ell^\varepsilon) \cap P_m) > 0$, it suffices to take $B_1 = C_\ell^\varepsilon$ and $B_2 = C_{\ell'}^\varepsilon$.

Now, for $\tilde{\omega} \in \chi^{-1}(B_1) \cap P_m$ it must be the case that $\chi(\tilde{\omega}) \in B_1$ and hence also that $\chi(T_k^{\ell m} \tilde{\omega}) \in B_1$ for all $\ell \in \mathbb{N}$. In particular, $T_k^{\ell m} \tilde{\omega} \in \chi^{-1}(B_1)$ for all $\ell \in \mathbb{N}$. However, $\chi^{-1}(B_1) \cap \chi^{-1}(B_2) = \emptyset$, so

$$\tilde{\mathbb{P}}(T_k^{\ell m}(\chi^{-1}(B_1) \cap P_m) \cap \chi^{-1}(B_2)) = 0 < \tilde{\mathbb{P}}(\chi^{-1}(B_1) \cap P_m)\mathbb{P}(\chi^{-1}(B_2)).$$

As T_k is mixing on $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty}, \tilde{\mathbb{P}})$ and $\{\ell m : \ell \in \mathbb{N}\}$ is a set of strictly positive upper density, this is a contradiction. Lemma 6 concludes the proof for this case, as

$$\begin{aligned} \mathbb{P}(X_{-t} = X_{-t}^k, \dots, X_t = X_t^k) &= \mathbb{P}(\chi(T_k^{-t}X(\omega)) = \chi(T^{-t}X(\omega)), \dots, \chi(T_k^tX(\omega)) = \chi(T^tX(\omega))) \\ &\geq \mathbb{P}(T_k^{-t}X(\omega) = T^{-t}X(\omega), \dots, T_k^tX(\omega) = T^tX(\omega)) \\ &= \tilde{\mathbb{P}}(T_k^{-t}\tilde{\omega} = T^{-t}\tilde{\omega}, \dots, T_k^t\tilde{\omega} = T^t\tilde{\omega}). \end{aligned}$$

Case 2: Approximation with non-ergodic processes. Let $T : (\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}}) \rightarrow (\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$ continue to be the left shift map associated with X . As in Case 1, we may approximate T by a sequence of mixing transformations $(T'_k)_{k \in \mathbb{N}}$ satisfying $T'_k \rightarrow T$ (uniform topology). We take the further step of approximating each T'_k with a transformation T_k which is not ergodic. As we have seen, the fact that $\tilde{\mathbb{P}}$ is non-atomic implies that there exist two disjoint sets $B_1, B_2 \subset \mathcal{X}$ such that $\tilde{\mathbb{P}}(\chi^{-1}(B_1)), \tilde{\mathbb{P}}(\chi^{-1}(B_2)) > 0$. We may pick a subset $A_k \subset \chi^{-1}(B_1) \sqcup \chi^{-1}(B_2)$ such that $\tilde{\mathbb{P}}(A_k \cap \chi^{-1}(B_1)), \tilde{\mathbb{P}}(A_k \cap \chi^{-1}(B_2)), \tilde{\mathbb{P}}(A_k^c \cap \chi^{-1}(B_1)), \tilde{\mathbb{P}}(A_k^c \cap \chi^{-1}(B_2)) > 0, 0 < \tilde{\mathbb{P}}(A_k) < 1/k$, and

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbf{1}_{\tilde{\omega} \in \chi^{-1}(B_1)} \mid \tilde{\omega} \in A_k] \equiv \tilde{\mathbb{P}}(A_k)^{-1} \int_{A_k} \mathbf{1}_{\tilde{\omega} \in \chi^{-1}(B_1)} d\tilde{\mathbb{P}} < \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbf{1}_{\tilde{\omega} \in \chi^{-1}(B_1)}].$$

T'_k induces a transformation on both A_k and A_k^c (cf. [41, §3.11]). In particular, for any $\tilde{\omega} \in \mathcal{X}^{\mathbb{Z}}$ and $B \in \mathfrak{B}^{\infty}$ let $t_{B,k}(\tilde{\omega}) = \min\{t > 0 : (T'_k)^t(\tilde{\omega}) \in B\}$. Then, the induced transformations $T_{A_k} : A_k \rightarrow A_k$ and $T_{A_k^c} : A_k^c \rightarrow A_k^c$ defined by

$$T_{A_k}(\tilde{\omega}) = (T'_k)^{t_{A_k,k}(\tilde{\omega})}(\tilde{\omega}), \quad T_{A_k^c}(\tilde{\omega}) = (T'_k)^{t_{A_k^c,k}(\tilde{\omega})}(\tilde{\omega})$$

are ergodic m.p. transformations on A_k and A_k^c , respectively. Define

$$T_k(\tilde{\omega}) \equiv \begin{cases} T_{A_k}(\tilde{\omega}) & \text{if } \tilde{\omega} \in A_k, \\ T_{A_k^c}(\tilde{\omega}) & \text{if } \tilde{\omega} \in A_k^c. \end{cases}$$

Then T_k is invertible, and for all $B \in \mathfrak{B}^{\infty}$

$$\tilde{\mathbb{P}}(T_k^{-1}(B)) = \tilde{\mathbb{P}}(T_{A_k}^{-1}(B \cap A_k) \cup T_{A_k^c}^{-1}(B \cap A_k^c)) = \tilde{\mathbb{P}}(B \cap A_k) + \tilde{\mathbb{P}}(B \cap A_k^c) = \tilde{\mathbb{P}}(B),$$

so $T_k \in M(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$. Note that T_k is not ergodic because it fixes the sets A_k and A_k^c . However, application of [41, Theorem 6.4.2 and Lemma 3.11.2] shows that T_k is weakly mixing on its restrictions to A_k and A_k^c , with respect to the conditional probabilities $\tilde{\mathbb{P}}|_{A_k}$ and $\tilde{\mathbb{P}}|_{A_k^c}$.

We again consider the stationary time series $X_t^k \equiv \chi(T_k^t \circ X(\omega))$ for $t \in \mathbb{Z}$, which has a standard probability distribution on $(\mathcal{X}^{\mathbb{Z}}, \mathfrak{B}^{\infty})$. As in the first case we must show that X^k is aperiodic. We again appeal to a proof by contradiction. Suppose that X^k is periodic with some positive probability, so that there is some $m \in \mathbb{N}$ and positive measure set $P_m = \{\tilde{\omega} : \chi(T_k^t\tilde{\omega}) = \chi(T_k^{t+m}\tilde{\omega}) \text{ for all } t \in \mathbb{Z}\}$; then $P_m \cap A_k$ or $P_m \cap A_k^c$ has positive measure. Suppose, without loss of generality, that the former is true (the same proof applies to both cases). Then, as in the preceding part, for all $\ell \in \mathbb{N}$ we have

$$T_k^{\ell m}(\chi^{-1}(B_1) \cap P_m \cap A_k) \cap (\chi^{-1}(B_2) \cap A_k) = \emptyset,$$

which is a contradiction to the fact that the restriction $T_k|_{A_k} = T_{A_k}$ is at least weakly mixing on A_k . Therefore (X^k) is aperiodic.

Now we argue that the time series X_t^k is not ergodic. By the ergodicity of T_{A_k} and the construction of A_k , for $\omega \in X^{-1}(A_k)$ we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t \mathbf{1}_{B_1}(X_j^k(\omega)) &= \lim_{n \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t \mathbf{1}_{X^{-1}(B_1)}(T_{A_k}^j X(\omega)) \stackrel{\text{a.s.}}{=} \mathbb{E}[\mathbf{1}_{\tilde{\omega} \in X^{-1}(B_1)} | \tilde{\omega} \in A_k] \\ &< \mathbb{E}[\mathbf{1}_{\tilde{\omega} \in X^{-1}(B_1)}] = \mathbb{P}(X_0 \in B_1). \end{aligned}$$

As $\mathbb{P}(X_0 \in B_1) = \mathbb{P}(X_0^k \in B_1)$, this implies that (X_t^k) is not ergodic.

Finally, we argue that $T_k \rightarrow T$ (uniform topology). By the definition of T_k we have $d(T'_k, T_k) = \tilde{\mathbb{P}}(T'_k \neq T_k) \leq 1 - \tilde{\mathbb{P}}(A_k^c \cap \{t_{A_k^c} = 1\})$. Moreover, $\{t_{A_k^c} = 1\} = (T'_k)^{-1}(A_k^c)$ and $\tilde{\mathbb{P}}((T'_k)^{-1}(A_k^c)) = \tilde{\mathbb{P}}(A_k^c) > 1 - k^{-1}$. So $\tilde{\mathbb{P}}(A_k^c \cap \{t_{A_k^c} = 1\}) > 1 - 2/k$, and we have

$$\liminf_{k \rightarrow \infty} d(T_k, T) \leq \liminf_{k \rightarrow \infty} d(T_k, T'_k) + \liminf_{k \rightarrow \infty} d(T'_k, T) = 0.$$

Lemma 6 then implies the desired convergence result, as in Case 1.

Case 3: Approximation with mean-ergodic processes. Every ergodic process is also mean-ergodic, so this follows directly from Case 1.

Case 4: Approximation with non-mean-ergodic processes. By Case 1 it suffices to consider the case where X is ergodic, so fix such a process and its law $\tilde{\mathbb{P}}$ on $\mathcal{X}^{\mathbb{Z}}$. Let $g = \chi$, the projection of $\tilde{\omega} \in \mathcal{X}^{\mathbb{Z}}$ onto its 0th coordinate. By assumption, $g \in L^1(\mathcal{X}^{\mathbb{Z}}, \tilde{\mathbb{P}})$. Because $\tilde{\mathbb{P}}$ is non-atomic, there exist disjoint sets $B_1, B_2 \subset \mathbb{R}$ such that $\tilde{\mathbb{P}}(\chi^{-1}(B_1)), \mathbb{P}(\chi^{-1}(B_2)) > 0$, $B_2 = B_1^c$, and $y \in B_1 \implies y > \mathbb{E}[X_0]$. As in the proof of Case 2, there exists a positive measure set $A_k \in \mathfrak{B}^{\infty}$ and a $\tilde{\mathbb{P}}$ -preserving transformation T_k such that $\tilde{\mathbb{P}}(A_k) \rightarrow 0$, $T_k \rightarrow T$ (uniform topology), and T_k fixes A_k and A_k^c and is measure-preserving and weakly mixing on each piece. Moreover, by choosing A_k so that $\tilde{\mathbb{P}}(A_k \cap B_2)$ is sufficiently small relative to $\tilde{\mathbb{P}}(A_k \cap B_1)$, we may clearly guarantee that $\mathbb{E}[\chi(\tilde{\omega}) | \tilde{\omega} \in A_k] > \mathbb{E}_{\tilde{\mathbb{P}}}[\chi(\tilde{\omega})] = \mathbb{E}[X_0^k]$.

Again letting $X_t^k(\omega) = \chi(T_t^k \circ X(\omega))$, we can quickly verify that (X^k) is stationary, aperiodic, and integrable by the method of Case 2. Hence, for all t , $\mathbb{E}[|X_t^k|] = \mathbb{E}[|X_0|] < \infty$. However, for $\omega \in A_k$, ergodicity of T_k on A_k implies

$$\lim_{n \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t X_j^k(\omega) = \lim_{n \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t \chi(T_k^j(X(\omega))) \stackrel{\text{a.s.}}{=} \mathbb{E}_{\tilde{\mathbb{P}}}[\chi(\tilde{\omega}) | \tilde{\omega} \in A_k] d\mathbb{P} > \mathbb{E}_{\tilde{\mathbb{P}}}[X_0(\tilde{\omega})] = \mathbb{E}[X_0^k],$$

which also implies that $(X_t^k)_{t \in \mathbb{Z}}$ is not mean-ergodic. As in Case 1, Lemma 6 concludes the proof. □

In the following lemma we use the convention from (2) of enriching sample spaces. In particular, given a random variable X with law \mathbb{P} supported on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, we identify X with the random variable $(X_t \circ \pi_1)_{t \in \mathbb{Z}}$ supported on the product space $(\Omega \times \Omega', \mathfrak{F} \times \mathfrak{F}', \mathbb{P} \times \mathbb{P}')$ and use \mathbb{P} to denote the product measure $\mathbb{P} \times \mathbb{P}'$.

Lemma 8. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and \mathcal{X} a Polish space satisfying $|\mathcal{X}| \geq 2$. Then, for any stationary \mathcal{X} -valued time series X on $(\Omega, \mathfrak{F}, \mathbb{P})$ and any standard non-atomic*

probability space $(\Omega', \mathfrak{F}', \mathbb{P}')$, there is a sequence (X^k) of stationary and aperiodic \mathcal{X} -valued time series on $(\Omega \times \Omega', \mathfrak{F} \otimes \mathfrak{F}', \mathbb{P} \times \mathbb{P}')$ satisfying (3) for all t . Moreover, if X is non-degenerate, (X^k) can be chosen to satisfy $X_0 \stackrel{d}{=} X_0^k$ for all $k \in \mathbb{N}$.

Proof. For all $t \in \mathbb{Z}$ let Y_t be i.i.d. copies of the random variable X_0 which are independent of X if X is non-degenerate, and otherwise i.i.d. copies of some non-degenerate \mathcal{X} -valued random variable. Let C_t^k be i.i.d. random variables satisfying

$$C_t^k = \begin{cases} 0 & \text{with probability } 1 - 1/k, \\ 1 & \text{with probability } 1/k, \end{cases}$$

chosen independently of all other random variables. The random vector (C^k, Y) can be constructed over the (standard) product probability space $\tilde{\Omega}' = \tilde{\Omega}^{\mathbb{Z}} \times (\mathcal{X}^{\mathbb{Z}})^{\mathbb{Z}}$ by invoking Lemma 5, where $\tilde{\Omega}$ is just the unit interval with Lebesgue measure and $\mathcal{X}^{\mathbb{Z}}$ is equipped with the pushforward measure induced by X . (C^k, Y) can thus also be constructed over the unit interval (say) using intervals instead of $\tilde{\Omega}'$, so by isomorphism it can be constructed over Ω' . Make this construction, so that (X, C^k, Y) has domain $\Omega \times \Omega'$ with appropriate product σ -algebra and measure $\mathbb{P} \times \mathbb{P}'$. According to our convention, let \mathbb{P} indicate this extension of our original measure to the product space. For all $t \in \mathbb{Z}$, $k \in \mathbb{N}$ let

$$X_t^k \equiv \begin{cases} X_t & \text{if } C_t = 0, \\ Y_t & \text{if } C_t = 1; \end{cases}$$

$X_t^k(\omega, \omega')$ can be written in the form $\gamma \circ (T \times S^1 \times S^2)^t \circ \xi$, where T is the m.p. shift of (10), S_1 and S_2 are m.p. left shifts on $\{0, 1\}^{\mathbb{Z}}$ and $\mathcal{X}^{\mathbb{Z}}$, respectively, and ξ is the map from $\Omega \times \Omega'$ to the product space $\mathcal{X}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}} \times \mathcal{X}^{\mathbb{Z}}$ with coordinates (X, C^k, Y) . As products of m.p. transformations are m.p., a standard proof along the lines of Proposition 3 implies that (X^k) is a stationary process, for all k . Clearly, (X^k) satisfies $X_0 \stackrel{d}{=} X_0^k$ in the non-degenerate case. Also, $\mathbb{P}(X_{-t} = X_{-t}^k, \dots, X_t = X_t^k) \geq \mathbb{P}(C_{-t} = \dots = C_t = 0) \geq 1 - (2t + 1)/k \rightarrow 1$. It remains only to show that (X^k) is aperiodic. This follows if we can show that, for all $m \in \mathbb{N}$, $\mathbb{P}(X_0^k = X_{\ell m}^k \text{ for all } \ell \in \mathbb{Z}) = 0$. Fix such an m . Let the random subset $S \subset \mathbb{Z}$ be given by $S \equiv \{t : C_t = 1\}$ and let A denote the subset of $\Omega \times \Omega'$ on which $|S \cap m\mathbb{Z}| = \infty$. Note that, by independence, $\mathbb{P}(A) = 1$, and

$$\begin{aligned} \mathbb{P}(X_0^k = X_{\ell m}^k \text{ for all } \ell \in \mathbb{Z}) &= \mathbb{P}(X_0^k = X_{\ell m}^k \text{ for all } \ell \in \mathbb{Z} \mid A) \\ &\leq \mathbb{P}(X_0^k = X_t^k \text{ for all } t \in S \cap m\mathbb{Z} \mid A) \\ &= \mathbb{P}(X_0^k = Y_t \text{ for all } t \in S \cap m\mathbb{Z} \mid A) \\ &= \mathbb{E}[\mathbb{P}(X_0^k = Y_t \text{ for all } t \in S \cap m\mathbb{Z} \mid X_0^k, S, A) \mid A] = 0, \end{aligned}$$

where the last line follows by the independence of Y from the other variables and the non-degeneracy of Y_t for all t . □

In conjunction with Proposition 4, Lemma 8 implies the approximation result of Theorem 1 (note that in Lemma 8, if X_0 is taken to be integrable, then so is X_0^k for all k). □

The proof of Theorem 2 is straightforward, and depends upon a connection between the conjugacy class of a left shift map associated with a time series and the set of codings of that time series.

Proof of Theorem 2. The theorem clearly follows if \mathcal{X} is a singleton, so suppose this is not the case. Lemma 8 implies that we may assume that X is aperiodic (write an element of $\Omega \times \Omega'$ as ω). By Lemmas 5 and 7, the probability spaces $(\mathcal{X}^{\mathbb{Z}}, X_{*}\mathbb{P})$ and $(\mathcal{Y}^{\mathbb{Z}}, Q)$ (with their respective Borel σ -algebras) are isomorphic. Let $\varphi: \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{Y}^{\mathbb{Z}}$ be an isomorphism between the measure spaces, and let $S: \mathcal{Y}^{\mathbb{Z}} \rightarrow \mathcal{Y}^{\mathbb{Z}}$ be the left shift.

The map $\varphi^{-1}S\varphi: \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}^{\mathbb{Z}}$ preserves $X_{*}\mathbb{P}$. Hence, [16, Theorem 7.14] implies that there is a sequence of m.p. transformations (η_k) each mapping $\mathcal{X}^{\mathbb{Z}}$ to $\mathcal{X}^{\mathbb{Z}}$ such that $\eta_k^{-1}\varphi^{-1}S\varphi\eta_k \rightarrow T$ (uniform topology), where T is the left shift map on $\mathcal{X}^{\mathbb{Z}}$. Define an \mathcal{X} -valued time series X^k by

$$X_t^k(\omega) = \chi((\eta_k^{-1}\varphi^{-1}S\varphi\eta_k)^t X(\omega)) = \chi(\eta_k^{-1}\varphi^{-1}S^t\varphi\eta_k X(\omega)),$$

where $\chi: \mathcal{X}^{\mathbb{Z}} \rightarrow \mathcal{X}$ is the projection onto the zeroth coordinate. Then, (3) holds by Lemma 6. Moreover, X^k is a coding of the time series $Y^k(\omega) = \varphi\eta_k X(\omega)$, which has law Q , where (4) holds with $\tilde{\chi} = \chi \circ \eta_k^{-1}\varphi^{-1}$. It is also clear that $X_0 \stackrel{d}{=} X_0^k$, which concludes the proof. \square

Proof of Theorem 3. By Rohlin's theorem [23], any aperiodic transformation T can be approximated in the uniform sense by strictly periodic transformations (see, e.g., [16, §7], and especially Corollary 7.12). As in Theorems 1 and 2, we can define $X_t^k(\omega) = \chi(T_k^t\omega)$, using Lemma 8 if X is not itself aperiodic. \square

5. Conclusion

By identifying Polish-space-valued time series with measure-preserving transformations and applying uniform approximation results from ergodic theory, we have demonstrated that time series can be approximated in a strong sense by other series having mixing and non-mixing characteristics, and codings of aperiodic processes. One of the corollaries of this approximation result is that the non-parametric ergodic hypothesis, among other hypotheses, is non-testable. There are also Baire category implications for mixing conditions.

It would be interesting to apply other approximation results from ergodic theory to time series. For instance, [16] also shows that a dense set of m.p. transformations can be embedded into measurable flows, which suggests an embedding of time series into continuous stochastic processes. There are also approximation results for non-singular transformations (see, e.g., [10]) that might be applied to non-stationary time series.

Acknowledgements

I wish to thank Joel Horowitz, Ivan Canay, Eric Auerbach, and seminar participants at Northwestern University for their helpful comments. I also thank the referees and the editors who handled this manuscript for their thoughtful and insightful remarks.

Funding information

There are no funding bodies to thank relating to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ADAMS, T. M. AND NOBEL, A. B. (1998). On density estimation from ergodic processes. *Ann. Prob.* **26**, 794–804.
- [2] AGEEV, O. N. AND SILVA, C. E. (2002). Genericity of rigid and multiply recurrent infinite measure-preserving and nonsingular transformations. *Topology Proc.* **26**, 357–365.
- [3] ALPERN, S. AND PRASAD, V. S. (1989). Coding a stationary process to one with prescribed marginals. *Ann. Prob.* **17**, 1658–1663.
- [4] ALPERN, S. AND PRASAD, V. S. (2008). Multitowers, conjugacies and codes: Three theorems in ergodic theory, one variation on Rokhlin’s lemma. *Proc. Am. Math. Soc.* **136**, 4373–4383.
- [5] BEZUGLYI, S., KWIATKOWSKI, J., and MEDYNETS, K. (2005). Approximation in ergodic theory, Borel and Cantor dynamics. *Contemp. Math.* **385**, 39–64.
- [6] BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley, Chichester.
- [7] BOGACHEV, V. I. (2007). *Measure Theory*, Vol. 2. Springer, Berlin.
- [8] CANAY, I. V., SANTOS, A. AND SHAIKH, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica* **81**, 2535–2559.
- [9] CARVALHO, S. L. AND CONDORI, A. (2021). Generic properties of invariant measures of full-shift systems over perfect Polish metric spaces. *Stoch. Dynam.* **21**, 2150040.
- [10] CHOKSI, J. AND PRASAD, V. (2006). Approximation and Baire category theorems in ergodic theory. In *Measure Theory and its Applications*, eds. J.-M. Belley, J. Dubois and P. Morales (Lect. Notes Math. **1033**). Springer, New York, pp. 94–113.
- [11] CORRADI, V., SWANSON, N. R. AND WHITE, H. (2000). Testing for stationarity–ergodicity and for comovements between nonlinear discrete time Markov processes. *J. Econometrics* **96**, 39–73.
- [12] COVER, T. M. AND THOMAS, J. A. (2012). *Elements of Information Theory*. Wiley, Chichester.
- [13] DOMOWITZ, I. AND EL-GAMAL, M. A. (1993). A consistent test of stationary–ergodicity. *Econometric Theory* **9**, 589–601.
- [14] DOMOWITZ, I. AND EL-GAMAL, M. A. (2001). A consistent nonparametric test of ergodicity for time series with applications. *J. Econometrics* **102**, 365–398.
- [15] EISNER, T. AND SERENY, A. (2009). Category theorems for stable semigroups. *Ergodic Theory Dynam. Syst.* **29**, 487–494.
- [16] FRIEDMAN, N. A. (1970). *Introduction to Ergodic Theory*. Van Nostrand Reinhold, New York.
- [17] GELFERT, K. AND KWIETNIAK, D. (2018). On density of ergodic measures and generic points. *Ergodic Theory Dynam. Syst.* **38**, 1745–1767.
- [18] GRAY, R. M., NEUHOFF, D. L. AND SHIELDS, P. C. (1975). A generalization of Ornstein’s \bar{d} distance with applications to information theory. *Ann. Prob.* **3**, 315–328.
- [19] GRAZZINI, J. (2012). Analysis of the emergent properties: Stationarity and ergodicity. *J. Artificial Soc. Social Sim.* **15**, 7.
- [20] GRILLENBERGER, C. AND KRENGEL, U. (1976). On marginal distributions and isomorphisms of stationary processes. *Math. Z.* **149**, 131–154.
- [21] GUERINI, M. AND MONETA, A. (2017). A method for agent-based models validation. *J. Economic Dynam. Control* **82**, 125–141.
- [22] HALMOS, P. R. (1944). In general a measure preserving transformation is mixing. *Ann. Math.* **45**, 786–792.
- [23] HALMOS, P. R. (2017). *Lectures on Ergodic Theory*. Dover Publications, Mineola, NY.
- [24] HANNEKE, S. (2021). Learning whenever learning is possible: Universal learning under general stochastic processes. *J. Mach. Learn. Res.* **22**, 1–116.
- [25] IOMMI, G., TODD, M. AND VELOZO, A. (2020). Upper semi-continuity of entropy in non-compact settings. *Math. Res. Lett.* **27**, 1055–1077.
- [26] ITÔ, K. (1984). *An Introduction to Probability Theory*. Cambridge University Press.
- [27] KALLENBERG, O. (2002). *Foundations of Modern Probability*. Springer, New York.
- [28] KARLIN, S. AND TAYLOR, H. E. (2012). *A First Course in Stochastic Processes*. Elsevier, Amsterdam.
- [29] KIEFFER, J. C. (1974). On the approximation of stationary measures by periodic and ergodic measures. *Ann. Prob.* **2**, 530–534.
- [30] Kieffer, J. C. (1980). On coding a stationary process to achieve a given marginal distribution. *Ann. Prob.* **8**, 131–141.

- [31] KIEFFER, J. C. (1983). On obtaining a stationary process isomorphic to a given process with a desired distribution. *Monatshefte Math.* **96**, 183–193.
- [32] LOCH, H., JANCZURA, J. AND WERON, A. (2016). Ergodicity testing using an analytical formula for a dynamical functional of alpha-stable autoregressive fractionally integrated moving average processes. *Phys. Rev. E* **93**, 043317.
- [33] ORNSTEIN, D. S. (1973). An application of ergodicity theory to probability theory. *Ann. Prob.* **1**, 43–58.
- [34] ORNSTEIN, D. S. AND WEISS, B. (1990). How sampling reveals a process. *Ann. Prob.* **18**, 905–930.
- [35] PARTHASARATHY, K. R. (1961). On the category of ergodic measures. *Illinois J. Math.* **5**, 648–656.
- [36] PLATT, D. (2020). A comparison of economic agent-based model calibration methods. *J. Economic Dynam. Control* **113**, 103859.
- [37] ROKHLIN, V. A. (1948). A general measure-preserving transformation is not mixing. *Dokl. Akad. Nauk* **60**, 349–351.
- [38] RÜSCHENDORF, L. AND SEI, T. (2012). On optimal stationary couplings between stationary processes. *Electron. J. Prob.* **17**, 1–20.
- [39] RYABKO, D. (2010). Discrimination between b -processes is impossible. *J. Theoret. Prob.* **23**, 565–575.
- [40] SHIELDS, P. C. (1996). *The Ergodic Theory of Discrete Sample Paths*. American Mathematical Society, Providence, RI.
- [41] SILVA, C. E. (2007). *Invitation to Ergodic Theory* (Student Math. Library **42**). American Mathematical Society, Providence, RI.
- [42] SRIVASTAVA, S. M. (1998). *A Course on Borel Sets*. Springer, New York.
- [43] TAO, T. (2012). *Topics in Random Matrix Theory*. American Mathematical Society, Providence, RI.
- [44] WALTERS, P. (2000). *An Introduction to Ergodic Theory*. Springer, New York.
- [45] WANG, H., WANG, C., ZHAO, Y. AND LIN, X. Toward practical approaches toward ergodicity analysis. *Theoret. Applied Climatology* **138**, 1435–1444.