STRICT IRREDUCIBILITY OF MARKOV CHAINS AND ERGODICITY OF SKEW PRODUCTS

PABLO LUMMERZHEIM, FELIX POGORZELSKI AND ELIAS ZIMMERMANN

Abstract. We consider a family of measure preserving transformations, which act on a common probability space and are chosen at random by a stationary ergodic Markov chain. This setting defines an instance of a random dynamical system (RDS), which may be described in terms of a step skew product. In many contexts it is desirable to know whether ergodicity of the family implies ergodicity of the skew product. Introducing the notion of strict irreducibility for Markov kernels we shall characterize the class of Markov chains for which the aforementioned implication holds true. We thereby extend a sufficient condition of Bufetov for the case of finite state Markov chains to general state spaces and show that it is in fact also necessary. As an application we obtain an explicit description of the limit in ergodic theorems for a suitable class of random transformations.

1 Introduction

Let $(E, \mathcal{E})$ and $(X, \mathcal{A})$ be measurable spaces. Fixing a probability space $(\Omega, \mathcal{C}, \nu)$ we shall consider a (measurable) family of transformations $(T_y)_{y \in E}$ on $X$, which are chosen randomly by a stationary stochastic process $(\xi_n)_{n=0}^\infty$ consisting of random variables $\xi_n: \Omega \to E$. This setting describes an instance of a random dynamical system (RDS). Random iterations of transformations play an important role in various areas of ergodic theory and related fields such as hyperbolic dynamics, perturbation theory, group actions or the theory of fractals, see e. g. [3], [15] or [22]. Throughout this paper we will always assume that the family of transformations $(T_y)_{y \in E}$ preserves a common probability measure $\mu$ on $X$. This situation arises for instance naturally in the study of measure preserving group actions. Let us illustrate this by giving some examples. In [29] Oseledets used random walks on groups of measure preserving transformations to obtain an ergodic theorem for general group actions. Aaronson and Lemańczyk studied iterations driven by random walks in connection with mixing properties of group actions on Poisson boundaries, cf. [2]. Random walks correspond to an i.i.d. selection of transformations. Iterations of measure preserving transformations, which are selected by a Markov chain, were considered by Grigorchuk in [17] and [18] to obtain an ergodic theorem for actions of the free group, which has been proved by different methods also in [4] and [27]. In a more general setting Markovian averages of random transformations were studied by Bufetov in [6], [7] and [8], leading to a general method for constructing ergodic theorems for group actions. Based on this method an ergodic theorem for a large class of Fuchsian groups was proved in [12], while general ergodic theorems for Markov groups were obtained in [10] and [11]. For further details and applications we refer to the surveys [9] and [15].

In what follows we shall suppose that the random variables $\xi_n$ are given in their canonical form, i. e. $(\Omega, \mathcal{C})$ is the product space $(E^\mathbb{N}_0, \mathcal{E}^\mathbb{N}_0)$ and $\xi_n$ is the projection to the $n$-th coordinate. In this setting stationarity (ergodicity) of the process $(\xi_n)_{n=0}^\infty$ amounts to the assumption that the measure $\nu$ is invariant (ergodic) with respect to the left shift $S$ on $\Omega$. An i.i.d. process is then obtained by taking $\nu$ to be the product measure $m^\mathbb{N}_0$ corresponding to some probability measure $m$ on $E$. On the other hand a
(stationary) Markov chain arises by considering a Markov measure $\nu$ given by a Markov kernel $\pi$:

$$E \xrightarrow[\pi]{} [0, 1]$$

and a $\pi$-invariant probability measure $m$ on $E$.

In many applications the ergodic behavior of the random iterations plays an important role. A fundamental tool for the study of this behavior is the step skew product $T$ on $\Omega \hat{\times} X$ with base $S$ and fiber maps $(T_y)_{y \in E}$. In the above setting it is easy to see that $T$ preserves the product measure $\nu \otimes \mu$. Given some function $f \in L^1(\mu)$ and denoting by $T_\omega^i$ the random composition $T_{\omega_{i-1}} \circ \ldots \circ T_{\omega_0}$, where we set $T_\omega^0 := I$ by convention, we may consider the random averages

$$A_nf(\omega, x) := \frac{1}{n} \sum_{i=0}^{n-1} f \circ T_\omega^i(x)$$

for any $(\omega, x) \in \Omega \times X$ and $n \in \mathbb{N}$. Observing that these averages are just the ordinary ergodic averages of the function $1 \otimes f$ with respect to the skew product $T$, we obtain that $A_nf$ converges $\nu \otimes \mu$-almost surely to some limit function $f^* \in L^1(\nu \otimes \mu)$ by Birkhoff’s ergodic theorem. The latter is given by the conditional expectation of $1 \otimes f$ under the $\sigma$-algebra $\Sigma_T$ of $T$-invariant sets. Assuming that the shift $S$ is ergodic, one might expect that the limit is independent of the $\Omega$-argument and equals almost surely the conditional expectation of $f$ under the $\sigma$-algebra $\Sigma$ of sets, which are invariant under the family $(T_y)_{y \in E}$. In order to give such a description of the limit function it is crucial to understand the $\sigma$-algebra $\Sigma_T$ of $T$-invariant sets. More precisely, considering the (trivial) $\sigma$-algebra $\Sigma_S$ of $S$-invariant sets, it would be desirable that for any family $(T_y)_{y \in E}$ of transformations the $\sigma$-algebra $\Sigma_T$ decomposes into the product of $\Sigma_S$ and $\Sigma$. As we shall see below, this is equivalent to the condition that the ergodicity of the family $(T_y)_{y \in E}$ implies the ergodicity of the skew product $T$.

Extending work of Pitt and Ulam-von Neumann, see [30] and [33], Kakutani showed in [20] that if the transformations are invertible and chosen independently by the same distribution, then the aforementioned implication holds true. The assumption of invertibility was later removed by Ryll-Nardzewski, see [31]. In the context of finite state Markov chains Bufetov introduced in [6], [7] and [8] the concept of strict irreducibility of transition matrices in order to give a sufficient condition for the above property to hold.

Our main goal in this paper is to characterize the class of general Markov chains showing the above property, namely that every skew extension with an ergodic family of transformations is ergodic. Introducing the notion of strict irreducibility for Markov kernels we shall extend Bufetov’s criterion from finite state spaces to arbitrary state spaces and show that it is also necessary for the property in question to hold. To this end we rely on the spectral theory of the Perron-Frobenius operator associated to a step skew product of the above type. More precisely, we will utilize a result of Kowalski, cf. [24], providing a convenient representation of the eigenfunctions corresponding to this operator. Using a characterization of strict irreducibility in terms of Markov operators this will allow us to obtain a particularly simple description of the invariant functions of the respective skew product, whenever the base shift arises from a strictly irreducible Markov kernel, cf. Theorem 4.2. In our main result, Theorem 4.3, we shall use this to show that a stationary Markov chain satisfies the aforementioned property if and only if the corresponding Markov kernel is strictly irreducible. As a consequence we are able to give an explicit description of the limit in ergodic theorems for the respective random dynamical systems. This generalizes Kakutani’s as well as Bufetov’s results.
The results of this work will appear also in the diploma thesis of one of the authors (P. L.), supervised by the other two authors (F. P.) and (E. Z.). This diploma project was pivotal to extend the setting under consideration from Markov shifts over countable state spaces to Markov shifts over arbitrary measurable state spaces.

The paper is organized as follows: Section 2 recalls some preliminaries on Markov operators and dynamical systems. In Section 3 we fix the setting and give a definition of strict irreducibility for Markov kernels. Section 4 is devoted to the proof of our main result, Theorem 4.3. Finally, in Section 5 we derive two versions of the ergodic theorem for random transformations, which are driven by Markov chains, see Theorems 5.1 and 5.2.

2 Preliminaries

In this section we collect some basic facts about Markov operators and dynamical systems. Let \((Y,\mathcal{B},\eta)\) be a probability space. By a measure preserving transformation we mean a measurable map \(T: Y \to Y\) such that \(\eta(T^{-1}D) = \eta(D)\) for all sets \(D \in \mathcal{B}\). The triple \((Y,\eta,T)\) is called a measure preserving dynamical system (MDS). A set \(D \in \mathcal{B}\) is invariant under \(T\) or \(T\)-invariant for short if it satisfies \(\eta(D\Delta T^{-1}D) = 0\). The collection of \(T\)-invariant sets forms a \(\sigma\)-algebra, which shall be denoted by \(\Sigma_T\). The transformation \(T\) is said to be ergodic if \(\Sigma_T\) is trivial, i. e. it contains only sets of trivial measure.

Given a bounded linear operator \(B: L^p(\eta) \to L^q(\eta)\), where \(p \in [1, \infty)\), we shall denote by \(B'\) the dual operator, i. e. the unique bounded linear operator \(B': L^q(\eta) \to L^p(\eta)\) such that

\[
\int_Y Bg \cdot h \, d\eta = \int_Y g \cdot B'h \, d\eta
\]

for all \(g \in L^p(\eta)\) and \(h \in L^q(\eta)\), where \(q\) denotes the dual exponent of \(p\). For \(p = 2\) we shall furthermore denote by \(B^*\) the adjoint operator, i. e. the unique bounded linear operator \(B^*: L^2(\eta) \to L^2(\eta)\) satisfying \(\langle Bg, h \rangle = \langle g, B^*h \rangle\) for all \(g \in L^2(\eta)\) and \(h \in L^2(\eta)\), where \(\langle \cdot, \cdot \rangle\) denotes the standard inner product of \(L^2(\eta)\). For any \(p \in [1, \infty]\) the Koopman operator \(\hat{T}: L^p(\eta) \to L^p(\eta)\) corresponding to some measure preserving transformation \(T\) on \(Y\) is defined by \(\hat{T}g := g \circ T\) for \(g \in L^p(\eta)\). It is easy to see that \(\hat{T}\) is linear and bounded. By a \(T\)-invariant function we mean a fixed function of \(T\). It is not difficult to verify that a function \(f \in L^p(\eta)\) is \(T\)-invariant if and only if it is \(\Sigma_T\)-measurable. Considering the Koopman operator \(\hat{T}\) on \(L^2(\eta)\) there is a unique bounded linear operator \(L_T: L^1(\eta) \to L^1(\eta)\) satisfying \(L_T^\ast = \hat{T}\), see [1 §1.3]. \(L_T\) is called the Perron-Frobenius operator corresponding to \(T\).

Given a measurable function \(f\) we shall write \(f \geq 0\) to indicate that \(\eta\)-almost everywhere \(f\) is real-valued and non-negative. A linear operator \(M: L^2(\eta) \to L^2(\eta)\) is called positive if \(f \geq 0\) implies \(Mf \geq 0\) for all \(f \in L^2(\eta)\). Positive operators are always bounded, cf. [14] Lemma 7.5. By a Markov operator we mean a positive operator \(M\) satisfying \(M1 = 1\) and

\[
\int_Y Mf \, d\eta = \int_Y f \, d\eta
\]

for all \(f \in L^2(\eta)\). Markov operators are contractions of unit operator norm and the class of Markov operators is closed under composition and taking adjoints, cf. [14] Section 13.1. We call a Markov operator \(M\) irreducible if \(M1_D = 1_D\) implies \(\eta(D) \in \{0, 1\}\) for all sets \(D \in \mathcal{B}\). This notion can be generalized to positive operators on arbitrary
Banach lattices, cf. [14, Chapter 7]. In the definition one may equivalently replace the condition $M1 \leq 1$ by the weaker condition $M1 \leq 1$.

3 Strict irreducibility of Markov kernels

In this section we develop the relevant background on Markov chains. Moreover, we introduce the notion of strict irreducibility for Markov kernels, which will play a key role in the subsequent sections. To this end let $(E, \mathcal{E})$ be a measurable space. A map $\pi: E \times \mathcal{E} \to [0, 1]$ is called a Markov kernel if for all $y \in E$ the map $\pi(y, \cdot)$ is a probability measure, which will sometimes be denoted by $\pi_y$ in the following, and for all $B \in \mathcal{E}$ the map $\pi(\cdot, B)$ is measurable. The product of two Markov kernels $\pi$ and $\kappa$ is given by

$$\pi\kappa(y, B) := \int_E \kappa(z, B) \, d\pi_y(z)$$

for $y \in E$ and $B \in \mathcal{E}$ and defines again a Markov kernel. We shall call a probability measure $m$ on $\mathcal{E}$ invariant under a Markov kernel $\pi$ or $\pi$-invariant for short if it satisfies

$$m(B) = \int_E \pi(y, B) \, dm(y)$$

for all $B \in \mathcal{E}$. It is not difficult to see that a probability measure $m$, which is invariant under two Markov kernels $\pi$ and $\kappa$, is also invariant under the product kernel $\pi\kappa$.

Given a Markov kernel $\pi$ together with a $\pi$-invariant probability measure $m$ we may define a bounded linear operator $P: L^2(m) \to L^2(m)$, which is a well known example of a Markov operator, by setting

$$Pf(y) := \int_E f(z) \, d\pi_y(z)$$

for $f \in L^2(m)$ and $y \in E$, see [13, Section 1.6]. The product of Markov kernels is compatible with the composition of the respective Markov operators in the following sense: If $P$ and $Q$ are the Markov operators corresponding to Markov kernels $\pi$ and $\kappa$ with joint invariant probability measure $m$, then the Markov operator corresponding to the product kernel $\pi\kappa$ is given by the composition $PQ$ of $P$ and $Q$.

Within this section we will frequently use the following notation: Given a set $B \in \mathcal{E}$ together with some exponent $n \in \mathbb{N}$ we shall denote by $U^n_B$ the set of all $y \in E$ such that $\pi^n(y, B) > 0$. Noting that $U^n_B$ is just the preimage of the interval $[0, 1]$ under the map $\pi^n(\cdot, B)$ we may conclude that $U^n_B$ is measurable for all $n \in \mathbb{N}$ as is the set

$$U_B := \bigcup_{n=1}^{\infty} U^n_B.$$
3.1 Proposition. A Markov kernel \( \pi \) is irreducible with respect to a \( \pi \)-invariant probability measure \( m \) if and only if the Markov operator \( P \) corresponding to \( \pi \) and \( m \) is irreducible.

**Proof:** Assume that \( P \) is not irreducible. Then there is some set \( B \in \mathcal{E} \) with \( m(B) \in (0,1) \) such that \( P\mathbb{1}_B = \mathbb{1}_B \). This implies that for \( m \)-almost all \( y \in B^c \) and every \( n \in \mathbb{N} \) we have

\[
\pi^n(y, B) = P^n \mathbb{1}_B(y) = \mathbb{1}_B(y) = 0.
\]

Since \( B \) and \( B^c \) have both positive measure, the latter violates the irreducibility of \( \pi \) with respect to \( m \).

To show the converse direction assume that \( P \) is irreducible and fix a set \( B \in \mathcal{E} \) with \( m(B) > 0 \). We have to show that \( m(U_B) = 1 \). Since \( m \) is \( \pi \)-invariant, we know that \( m(U_B) > 0 \). Furthermore we have \( P\mathbb{1}_{U_B}(y) = \pi(y, U_B) = 0 \) for \( m \)-almost all \( y \in U_B \) and \( P\mathbb{1}_{U_B}(y) = \pi(y, U_B) = 1 \) for \( m \)-almost all \( y \in U_B \), which in sum gives \( P\mathbb{1}_{U_B} = \mathbb{1}_{U_B} \). By the irreducibility of \( P \) this implies that \( U_B \) has trivial measure. Since we already know that \( m(U_B) > 0 \), we obtain \( m(U_B) = 1 \). \( \square \)

The above proposition shows that the notion of irreducibility of Markov kernels introduced before is tightly related to the notion of \( m \)-irreducibility in the theory of Markov chains, cf. [28, Section 2.2] and [13, Section 9.2]. However, the latter is slightly stronger in the sense that it requires that for any \( B \in \mathcal{E} \) with \( m(B) > 0 \) all \( y \in E \) admit some \( n \in \mathbb{N} \) (which may depend on \( y \)) such that \( \pi^n(y, B) > 0 \). It will turn out that from a dynamical point of view the weaker notion of irreducibility introduced above is more convenient. It is for instance equivalent to the ergodicity of the Markov shift corresponding to \( \pi \) and \( m \) as we shall see in Proposition 3.2 below.

In what follows we shall denote by \((\Omega, \mathcal{E})\) the product space \((E^\mathbb{N}, \mathcal{E}^\mathbb{N})\). Consider a Markov kernel \( \pi \) and a probability measure \( m \) as above. Then there is a unique probability measure \( \nu \) on \( \mathcal{E} \) satisfying

\[
\nu(B_0 \times \cdots \times B_{k-1} \times \Omega) = \int_{B_0} \cdots \int_{B_{k-1}} d\pi_{y_{k-2}}(y_{k-1}) \cdots d\pi_{y_0}(y_1) dm(y_0)
\]
for every $k \in \mathbb{N}$ and all sets $B_0, \ldots, B_{k-1} \in \mathcal{E}$, cf. [13, Theorem 3.1.2]. This is essentially a consequence of Carathéodory’s extension theorem. The measure $\nu$ is called the Markov measure corresponding to $\pi$ and $m$. Consider the shift map $S : \Omega \to \Omega$ given by

$$S(\omega_0 \omega_1 \ldots) := \omega_1 \omega_2 \ldots$$

for $\omega \in \Omega$. It is not difficult to see that $S$ is $\mathcal{E}$-measurable. We shall call the Markov measure $\nu$ stationary if $S$ is $\nu$-preserving. It is easy to check that $\nu$ is stationary if and only if the measure $m$ is $\pi$-invariant. In this case we obtain an MDS $(\Omega, \nu, S)$, which we shall call a Markov shift. A Bernoulli shift is obtained by considering the trivial kernel $\pi$ with respect to a probability measure $m$ given by $\pi_y := m$ for all $y \in E$, in which case $\nu$ coincides with the product measure $m^\infty$. The next proposition connects the ergodicity of the Markov shift $S$ to the irreducibility of the kernel $\pi$ with respect to the measure $m$. Noting that the trivial kernel is irreducible this contains the well known fact that Bernoulli shifts are ergodic as a special case.

3.2 Proposition. Let $\pi$ be a Markov kernel with invariant probability measure $m$. Then the respective Markov shift $S$ is ergodic if and only if $\pi$ is irreducible with respect to $m$.

Proof: Recall that a set $B \in \mathcal{E}$ is called absorbing if we have $\pi(y, B) = 1$ for all $y \in B$. It is well known that the Markov shift $S$ corresponding to $\pi$ and $m$ is ergodic if and only if $m(B) \in \{0, 1\}$ for every absorbing set $B \in \mathcal{E}$, see [13, Theorem 5.2.1].

Assume that $S$ is not ergodic. Then there is an absorbing set $B \in \mathcal{E}$ with $m(B) \in (0, 1)$. We claim that $\pi^n(y, B^c) = 0$ for all $y \in B$ and $n \in \mathbb{N}$. This can be shown inductively as follows: For $n = 1$ the claim follows from the definition of an absorbing set. Now for $n > 1$ we have

$$\pi^n(y, B^c) = \int_B \pi(z, B^c) \, d\pi^{n-1}_y(z) + \int_{B^c} \pi(z, B^c) \, d\pi^{n-1}_y(z)$$

for all $y \in B$. The function integrated in the first term vanishes on $B$ by the fact that $B$ is absorbing. Furthermore, by the induction hypothesis, we have $\pi^{n-1}(y, B^c) = 0$ for all $y \in B$, so we integrate over a null set in the second term. Therefore both terms are zero, which gives $\pi^n(y, B^c) = 0$ for all $y \in B$. Since $B$ and $B^c$ have positive measure by assumption, we obtain that $\pi$ is not irreducible with respect to $m$.

Conversely assume that $\pi$ is not irreducible with respect to $m$. Then there is a set $B \in \mathcal{E}$ with $m(B) > 0$ such that $U^c_B$ has positive measure. Since $m$ is $\pi$-invariant, $U_B$ has also positive measure, so we get $m(U^c_B) \in (0, 1)$. Recall that we have $\pi(y, U_B) = 0$ and thus $\pi(y, U^c_B) = 1$ for all $y \in U^c_B$, so $U^c_B$ is a non trivial absorbing set. This shows that $S$ is not ergodic.

We shall now introduce the central notion of this paper. Given a Markov kernel $\pi$ admitting a $\pi$-invariant probability measure $m$ we will call a set $B \in \mathcal{E}$ deterministic if for $m$-almost all $y \in E$ we have either $\pi(y, B) = 0$ or $\pi(y, B) = 1$. Based on this we shall say that $\pi$ is strictly irreducible with respect to $m$ if for all deterministic sets $B \in \mathcal{E}$ we have $m(B) \in \{0, 1\}$. As we shall see below, the latter condition can also be expressed in terms of the Markov operator $P$ corresponding to $\pi$ and $m$. To this end we will rely on the following general property of irreducible Markov operators.

3.3 Lemma. Let $(Y, \eta)$ be a probability space and $M : L^2(\eta) \to L^2(\eta)$ be an irreducible Markov operator. Consider functions $g, h \in L^2(\eta)$ with $g, h \geq 0$ such that $g + h = 1$ and $\langle Mg, h \rangle = 0$. Then either $g = 0$ or $h = 0$. 


Ergodicity of skew products

Proof: Let $D$ denote the preimage of the interval $(0, \infty)$ under the function $h$. Then $D$ is obviously measurable. Furthermore, since $Mg \geq 0$, $\langle Mg, h \rangle = 0$ implies that $\mathbb{1}_D Mg = 0$. On the other hand we have

$$Mg + Mh = M(g + h) = M\mathbb{1} = \mathbb{1},$$

so we obtain

$$\mathbb{1}_DMh = \mathbb{1}_DMg + \mathbb{1}_DMh = \mathbb{1}_D(Mg + Mh) = \mathbb{1}_D.$$ 

Both together yields

$$\int_D h \, d\eta = \int_Y h \, d\eta = \int_Y Mh \, d\eta \geq \int_D Mh \, d\eta$$

$$= \int_D \mathbb{1} \, d\eta = \eta(D).$$

The assumptions on $g$ and $h$ imply that $h \leq \mathbb{1}$, so we may conclude that $h = \mathbb{1}_D$. This gives $g = \mathbb{1}_{D^c}$ and thus

$$\mathbb{1}_D M \mathbb{1}_{D^c} = \mathbb{1}_D Mg = 0.$$

Using that $Mh \geq 0$ we obtain

$$M \mathbb{1}_{D^c} = Mg = \mathbb{1} - Mh \leq \mathbb{1}.$$ 

Both together implies

$$M \mathbb{1}_{D^c} = \mathbb{1}_D M \mathbb{1}_{D^c} + \mathbb{1}_D M \mathbb{1}_{D^c} \leq \mathbb{1}_{D^c}.$$ 

By the irreducibility of $M$ it follows that $D^c$ and thus $D$ has trivial measure. Accordingly we obtain that either $g = 0$ or $h = 0$. 

The above lemma allows us to characterize the property of strict irreducibility in terms of the irreducibility of Markov operators as follows.

3.4 Proposition. Let $\pi$ be a Markov kernel and $m$ be a $\pi$-invariant probability measure. Let $P$ denote the Markov operator corresponding to $\pi$ and $m$. Then $P^*P$ is irreducible if and only if $PP^*$ is irreducible. Both is the case if and only if $\pi$ is strictly irreducible with respect to $m$.

Proof: To show the first equivalence assume that $P^*P$ is irreducible and fix a set $B \in \mathcal{E}$ with $PP^*\mathbb{1}_B = \mathbb{1}_B$. Then we have

$$\langle P^* \mathbb{1}_B, P^* \mathbb{1}_{B^c} \rangle = \langle PP^* \mathbb{1}_B, \mathbb{1}_{B^c} \rangle = \langle \mathbb{1}_B, \mathbb{1}_{B^c} \rangle = 0.$$

Since $P^* \mathbb{1}_B \geq 0$ and $P^* \mathbb{1}_{B^c} \geq 0$ this implies $PP^* \mathbb{1}_B P^* \mathbb{1}_{B^c} = 0$. Accordingly, using that $P^* \mathbb{1}_B + P^* \mathbb{1}_{B^c} = P^* \mathbb{1} = \mathbb{1}$, we obtain $P^* \mathbb{1}_B = \mathbb{1}_C$ and $P^* \mathbb{1}_{B^c} = \mathbb{1}_{C^c}$ for some set $C \in \mathcal{E}$. It is easy to see that we also have $PP^* \mathbb{1}_{B^c} = \mathbb{1}_{B^c}$ by the fact that $PP^*$ is a Markov operator. Thus we get

$$\langle P^* P \mathbb{1}_C, \mathbb{1}_{C^c} \rangle = \langle P^* PP^* \mathbb{1}_B, P^* \mathbb{1}_{B^c} \rangle$$

$$= \langle PP^* \mathbb{1}_B, PP^* \mathbb{1}_{B^c} \rangle$$

$$= \langle \mathbb{1}_B, \mathbb{1}_{B^c} \rangle = 0.$$
Since $P^*P$ is irreducible by assumption, we may apply Lemma 3.3 to obtain that $m(C) \in \{0, 1\}$. If $m(C) = 1$ this gives $P^*1_B = 1$. However, since $P^*$ is bounded with $\|P^*\| = |P| = 1$, this is only possible if $m(B) = 1$. In the case $m(C) = 0$ we may use the same argument to obtain that $m(B^c) = 1$. In sum we see that $B$ has trivial measure. This shows that $PP^*$ is irreducible. A symmetric argument gives the reverse implication.

To show the second equivalence assume again that $P^*P$ is irreducible. Let $B \in \mathcal{E}$ be a deterministic set. Then we have $\pi(y,B)\pi(y,B^c) = 0$ for $m$-almost all $y \in E$ and thus

$$\langle P^*P1_B, 1_{B^c} \rangle = \langle P1_B, P1_{B^c} \rangle = \int_E \pi(y,B)\pi(y,B^c) \, dm(y) = 0,$$

so by Lemma 3.3 we obtain $m(B) \in \{0, 1\}$. This shows that $\pi$ is strictly irreducible with respect to $m$. To show the reverse direction assume that $\pi$ is strictly irreducible with respect to $m$ and fix a set $B \in \mathcal{E}$ with $P^*P1_B = 1_B$. Then we have

$$\int_E \pi(y,B)\pi(y,B^c) \, dm(y) = \langle P1_B, P1_{B^c} \rangle = \langle P^*P1_B, 1_{B^c} \rangle = \langle 1_B, 1_{B^c} \rangle = 0.$$

This implies $\pi(y,B)\pi(y,B^c) = 0$ and thus $\pi(y,B) \in \{0, 1\}$ for $m$-almost all $y \in E$. Accordingly $B$ is deterministic, so we get $m(B) \in \{0, 1\}$. This proves that $P^*P$ is irreducible.

Note that in the proof of the first statement of Proposition 3.4 we did not use any specific properties of the operator $P$. In fact the equivalence claimed in this statement holds for arbitrary Markov operators on $L^2$-spaces.

Let us remark that in the case that $(E, \mathcal{E})$ is a standard Borel space the adjoint operator $P^*$ of $P$ admits a more concrete description, which gives rise to a further characterization of strict irreducibility. This is due to the fact that in this setting there exists a Markov kernel $\pi^*$ such that

$$\int_E \int_E f(y,z) \, d\pi_y(z) \, dm(y) = \int_E \int_E f(y,z) \, d\pi^*_z(y) \, dm(z)$$

for all $f \in L^2(m)$. Such a kernel is called a reverse kernel of $\pi$. Indeed, the kernels $\pi$ and $\pi^*$ coincide with the regular conditional distributions of the unique measure $\eta$ on $\mathcal{E} \otimes \mathcal{E}$ satisfying

$$\eta(B_0 \times B_1) = \int_{B_0} \pi(y,B_1) \, dm(y)$$

for all $B_0, B_1 \in \mathcal{E}$, which exist by the disintegration theorem for measures whenever $(E,\mathcal{E})$ is a standard Borel space, cf. [21, Chapter 17]. The disintegration theorem guarantees moreover that the reverse kernel $\pi^*$ is unique up to $m$-null sets, i.e. for any other reverse kernel $\kappa$ one has $\kappa_y = \pi^*_y$ for $m$-almost all $y \in E$.

It is readily seen that $m$ is also $\pi^*$-invariant. Denote by $P^*$ the Markov operator corresponding to $\pi^*$ and $m$. By the essential uniqueness of the reverse kernel $P^*$ is
Ergodicity of skew products

unique as a Markov operator on $L^2(m)$. As an immediate consequence of the defining equation of the reverse kernel one obtains then

$$\langle Pg, h \rangle = \langle g, P^*h \rangle$$

for all $g, h \in L^2(m)$. This implies that $P^*$ coincides indeed with the adjoint operator of $P$. By Proposition 3.3 and the foregoing results this shows that $\pi^*\pi$ is irreducible with respect to $m$ if and only if $\pi\pi^*$ is irreducible with respect to $m$ and both is equivalent to the strict irreducibility of $\pi$ with respect to $m$.

The next proposition shows that strict irreducibility is indeed a stronger property than irreducibility. In fact, using Lemma 3.3, it is not difficult to give an abstract argument for this fact on the level of Markov operators. However, the proof of the following proposition, which operates on the level of Markov kernels, might be more instructive in the present setting.

3.5 Proposition. A Markov kernel $\pi$ which is strictly irreducible with respect to an invariant probability measure $m$ is also irreducible with respect to $m$.

Proof: Assume that $\pi$ is not irreducible with respect to $m$. Then there is some set $B \in \mathcal{E}$ with $m(B) > 0$ such that $m(U_B) < 1$. However, as noted several times already, we have $m(U_B) > 0$ by the $\pi$-invariance of $m$ as well as $\pi(y, U_B) = 0$ for all $y \in U_B$ and $\pi(y, U_B) = 1$ for $m$-almost all $y \in U_B$. This shows that $U_B$ is a non-trivial deterministic set. Thus $\pi$ is not strictly irreducible with respect to $m$.

In the remainder of this section we shall briefly discuss the notion of strict irreducibility in the context of discrete state spaces, which amounts to the assumption that $E$ is a countable set. In this case a probability measure corresponds to a map $m : E \rightarrow [0, 1]$ whose values sum up to 1, while a Markov kernel may be identified with a map $\pi : E \times E \rightarrow [0, 1]$ such that the functions $\pi(i, \cdot)$ define probability measures for all $i \in E$. The product of two Markov kernels $\pi$ and $\kappa$ is then given by the formula

$$\pi\kappa(i, j) = \sum_{k \in E} \pi(i, k)\kappa(k, j)$$

for any $i, j \in E$. Furthermore, the invariance of a probability measure $m$ with respect to a Markov kernel $\pi$ is equivalent to the validity of the equation

$$\sum_{j \in E} m(j)\pi(j, i) = m(i)$$

for all $i \in E$. In the following we confine ourselves to probability measures $m$, which are strictly positive in the sense that $m(i) > 0$ for all $i \in E$. We shall always assume that $\pi$ admits a strictly positive invariant measure $m$. Note that this assumption is just a convenience, since by the invariance condition states with probability zero will never be reached, so in general we may just restrict $\pi$ and $m$ to those states, which have positive initial probability.

Under the above assumption the irreducibility of $\pi$ with regard to a $\pi$-invariant probability measure $m$ does not depend on $m$ and is equivalent to the fact that for all $i, j \in E$ there is some $n \in \mathbb{N}$ such that $\pi^n(i, j) > 0$. So it coincides with the usual notion of irreducibility in the theory of Markov chains with discrete state space. Moreover, the reverse kernel $\pi^*$ of $\pi$ is explicitly given by the formula

$$\pi^*(i, j) := \frac{m(j)\pi(j, i)}{m(i)}$$
for \( i, j \in E \). In this setting there are two convenient ways to characterize the class of strictly irreducible Markov kernels, which shall be described next.

Consider the relation \( \sim \) on \( E \), which takes place between two states \( i \) and \( j \) if there is some state \( k \in E \) such that \( \pi(k, i) > 0 \) and \( \pi(k, j) > 0 \). Obviously, this defines a symmetric relation. Let \( \sim^* \) denote the transitive closure of \( \sim \), i.e., we have \( i \sim^* j \) if and only if \( i = j \) or \( i \sim k_1, k_1 \sim k_2, \ldots, k_r \sim j \) for some sequence of states \( k_1, \ldots, k_r \in E \). Then \( \sim \) defines an equivalence relation. It is not difficult to see that

the graph \( (E, \sim) \) is connected if and only if the equivalence relation \( \sim \) has only one equivalence class. Moreover, using the explicit formula for the reverse Markov kernel from above, a straightforward computation shows that both conditions are equivalent to the irreducibility of the kernel \( \pi^* \pi \). In fact, observing that from any state one may reach exactly one equivalence class of \( \sim \), one can show that in this setting the non-empty deterministic sets are precisely the unions of equivalence classes of \( \sim \). In particular, the property of strict irreducibility is an intrinsic property of the kernel \( \pi \), which does not depend on the choice of the \( \pi \)-invariant measure \( m \).

A second characterization of strict irreducibility can be obtained by considering the dual relation \( \sim^* \), which takes place between two states \( i, j \in E \) if there is a state \( k \in E \) such that \( \pi(i, k) > 0 \) and \( \pi(j, k) > 0 \). Using similar arguments as before one can show that the connectedness of the graph \( (E, \sim^*) \) is equivalent to the fact that the transitive closure \( \sim^* \) of \( \sim^* \) has only one equivalence class, which in turn is equivalent to the irreducibility of the kernel \( \pi \pi^* \). However, from the discussion following Proposition 3.4 we know that \( \pi^* \pi \) is irreducible if and only if \( \pi \pi^* \) is irreducible, which can also be directly verified using the graph theoretic characterizations from above and the fact that strict irreducibility implies irreducibility by Proposition 3.5. Thus we have obtained two purely combinatorial characterizations of strict irreducibility.

A connection of strict irreducibility to the triviality of the symmetric \( \sigma \)-algebra of a Markov chain is due to Grigorenko. In [19] he shows that in the discrete setting the latter is equivalent to the combinatorial characterization of strict irreducibility just given. This gives a generalization of the well known Hewitt-Savage theorem to a certain class of Markov chains. Grigorenko’s result was extended to Harris Markov chains with general state space by Bezhaeva and Oseledets in [5].

The notion of strict irreducibility was introduced by Bufetov in the case of a finite state space, see [6], [7] and [8], where it can be stated as a property of stochastic matrices. In this setting a Markov kernel can be described by a row stochastic matrix \( \Pi \), while an invariant probability measure may be identified with an invariant probability vector of \( \Pi \). The product of two Markov kernels corresponds to the usual matrix product. Let us assume that \( \Pi \) admits a strictly positive invariant probability vector. Then by the above discussion we have the following. The irreducibility of the Markov kernel does not depend on the particular choice of the invariant vector and coincides with the irreducibility of the matrix \( \Pi \) in the sense that for some \( n \in \mathbb{N} \) the sum \( \Pi + \Pi^2 + \cdots + \Pi^n \) has only positive entries. It is not difficult to verify that the strict irreducibility of the Markov kernel, which is also independent of the choice of the invariant vector, is equivalent to the irreducibility of the matrix \( \Pi^T \Pi \). The latter is in turn equivalent to the irreducibility of the matrix \( \Pi \Pi^T \) and implies the irreducibility of \( \Pi \). In [7] and [8] strict irreducibility is defined using the latter condition. In [6] the terminology of strongly and \( * \)-strongly connected Markov measures is used to refer to conditions, which are equivalent to the two combinatorial characterizations of strict irreducibility given above. By the foregoing discussion all of these notions describe equivalent concepts.
The authors have been informed that in the upcoming preprint [32] of Tserunyan and Zomback a characterization of weak mixing for the boundary action of finitely generated free semigroups with respect to Markov measures was obtained, showing in particular that it is equivalent to the strict irreducibility of the transition matrix.

4 Ergodicity of step skew products

In this section we shall use the notion of strict irreducibility to characterize the class of Markov shifts with the property that every skew extension with an ergodic family of transformations is ergodic. The characterization we shall give extends a criterion of Bufetov for Markov shifts with finite state space to general state spaces and shows that the respective condition is not only sufficient but also necessary for the above property to hold. In fact, it will turn out that this class consists precisely of those Markov shifts arising from strictly irreducible Markov kernels. Noting that trivial kernels are strictly irreducible with respect to their invariant probability measures, this includes in particular all Bernoulli shifts. It thus generalizes Bufetov’s as well as Kakutani’s results, cf. [6, Theorem 5] and [20, Theorem 3].

Let \((E, \mathcal{E})\) be a measurable space. We call a family \((T_y)_{y \in E}\) of transformations \(T_y\) on a further measurable space \((X, \mathcal{A})\) measurable if the map \((y, x) \mapsto T_y(x)\) is measurable with respect to \(\mathcal{E} \otimes \mathcal{A}\) and \(\mathcal{A}\). Considering a Markov shift \((\Omega, \nu, S)\) with state space \(E\) such a family gives rise to a measurable skew product \(T\) on \(\Omega \otimes X\) defined by

\[
T(\omega, x) := (S\omega, T_{\omega_0}x)
\]

for \((\omega, x) \in \Omega \times X\), which is an instance of so called step skew product. The mapping \((n, \omega, x) \mapsto T^n_\omega(x)\), where \(T^n_\omega\) denotes the random composition \(T_{\omega_{n-1}} \circ \cdots \circ T_{\omega_0}\) for \(n \in \mathbb{N}_0\), defines a (measurable) random dynamical system in the sense of [3]. In what follows we will always assume that the transformations in the family \((T_y)_{y \in E}\) preserve a common probability measure \(\mu\) on \(X\). It is easily verified that in this case \(T\) preserves the product measure \(\nu \otimes \mu\).

It will turn out that the invariant functions of \(T\) are of a particular simple form. This is a consequence of a result of Kowalski concerning the eigenfunctions of the Perron-Frobenius operator associated to step skew product as above, which may be stated as follows. Let \(L_T\) denote the Perron-Frobenius operator of \(T\) and consider a function \(\varphi \in L^1(\nu \otimes \mu)\) with \(L_T \varphi = \lambda \varphi\) for some \(\lambda \in \mathbb{C}\) with \(|\lambda| = 1\). Then by [24, Theorem 3.1] there is a function \(\hat{\varphi} \in L^1(\mu \otimes \mu)\) satisfying

\[
\varphi(\omega, x) = \hat{\varphi}(\omega_0, x)
\]

for \(\nu \otimes \mu\)-almost all \((\omega, x) \in \Omega \times X\). In other words the eigenfunctions of \(L_T\) depend only on the first entry of the \(\Omega\)-argument. This property transfers to \(T\)-invariant functions as follows.

4.1 Proposition. Let \((\Omega, \nu, S)\) be a Markov shift with state space \(E\) and \((T_y)_{y \in E}\) be a measurable family of measure preserving transformations on a probability space \((X, \mu)\). Let \(T\) be the arising step skew product and \(\varphi \in L^1(\nu \otimes \mu)\) be a \(T\)-invariant function. Then there is a function \(\hat{\varphi} \in L^1(\mu \otimes \mu)\) such that for \(\nu \otimes \mu\)-almost all \((\omega, x) \in \Omega \times X\) we have

\[
\varphi(\omega, x) = \hat{\varphi}(\omega_0, x).
\]
Proof: Consider the Koopman operator \( \hat{T} \) on \( L^\infty(\nu \otimes \mu) \) corresponding to \( T \) and let \( \varphi \in L^1(\nu \otimes \mu) \) be a \( T \)-invariant function. Then by the \( T \)-invariance of the measure \( \nu \otimes \mu \) we obtain
\[
\int_{\Omega \times X} \mathcal{L}_T \varphi \cdot \psi \, d\nu \otimes \mu = \int_{\Omega \times X} \varphi \cdot \hat{T} \psi \, d\nu \otimes \mu \\
= \int_{\Omega \times X} \varphi \circ T \cdot \psi \circ T \, d\nu \otimes \mu \\
= \int_{\Omega \times X} \varphi \cdot \psi \, d\nu \otimes \mu
\]
for all \( \psi \in L^\infty(\nu \otimes \mu) \). This shows that \( \mathcal{L}_T \varphi = \varphi \), so the claimed representation of \( \varphi \) follows from Kowalski’s theorem.

Let \( \Sigma \) denote the \( \sigma \)-algebra of sets \( A \in \mathcal{A} \) such that up to a \( \mu \)-null set we have \( T_y^{-1}(A) = A \) for \( m \)-almost all \( y \in E \). We shall call the family \( (T_y)_{y \in E} \) ergodic if \( \Sigma \) is trivial. It is not difficult to see that the ergodicity of the skew product \( T \) implies the ergodicity of the family \( (T_y)_{y \in E} \) as well as the ergodicity of the Markov shift \( S \). However, the converse does not hold in general.

To see this consider a 3-point set \( X := \{1, 2, 3\} \) equipped with the uniform probability measure \( \mu \) giving every point equal weight. Consider furthermore the Markov chain with state space \( \{0, 1\} \) corresponding to the transition matrix
\[
\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\]
and the \( \Pi \)-invariant initial probability vector \( (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T \). Then \( \Pi \) is irreducible, so the arising Markov shift \( (\{0, 1\}^{\mathbb{N}_0}, \mathcal{F}, S) \) is ergodic. Let \( T_0 \) be a non-trivial permutation of \( X \) and set \( T_1 := T_0^{-1} \). Then \( T_0 \) and \( T_1 \) are measure preserving and ergodic with respect to \( \mu \), in particular the family \( \{T_0, T_1\} \) is ergodic. To see that the arising step skew product \( T \) is not ergodic observe that the measure \( \mu \) is concentrated on the sequences \( \omega^0 := 010101... \) and \( \omega^1 := 101010... \) with equal weight \( \frac{1}{2} \). Moreover, it is not difficult to check that the set \( \{ (\omega^0, 1), (\omega^1, 0) \} \) is \( T \)-invariant. However, this set has measure \( \frac{1}{3} \). So \( T \) admits a non-trivial invariant set and can thus not be ergodic.

In Theorem 4.3 below we shall see that the converse implication holds true if the Markov shift arises from a strictly irreducible Markov kernel (in fact, it is easy to see that the matrix \( \Pi \) in the example above is not strictly irreducible). In order to obtain such a result we have to improve Proposition 4.1.

Let \( \pi \) be a Markov kernel. Fixing \( y \in E \) and denoting by \( \delta_y \) the Dirac measure concentrated in \( y \) we may consider the Markov measure \( \nu_y \) arising from \( \delta_y \) and \( \pi \). By definition the measure \( \nu_y \) takes the form
\[
\nu_y(B_0 \times \cdots \times B_{k-1} \times \Omega) = \mathbb{1}_{B_0}(y) \int_{B_1} \cdots \int_{B_{k-1}} d\pi_{z_{k-2}}(z_{k-1}) \cdots d\pi_y(z_1) \tag{4.1}
\]
for all sets \( B_0, \ldots, B_{k-1} \in \mathcal{E} \) and every \( k \in \mathbb{N} \). Based on this a standard approximation argument shows that for any \( k \in \mathbb{N} \) and any bounded measurable function \( h: E^k \to \mathbb{R} \) we obtain
\[
\int_{\Omega} h(\omega_0, \ldots, \omega_{k-1}) \, d\nu_y(\omega) = \int_E \cdots \int_E h(y, z_1, \ldots, z_{k-1}) \, d\pi_{z_{k-2}}(z_{k-1}) \cdots d\pi_y(z_1). \tag{4.2}
\]
Ergodicity of skew products

One can show that the family of measures \((\nu_y)_{y \in E}\) provides a disintegration of the measure \(\nu\) with respect to \(m\), i.e. for all sets \(C \in \mathcal{E}\) we have

\[
\nu(C) = \int_E \nu_y(C) \, dm(y),
\]

see [13, Proposition 3.1.3].

We are now able to improve the representation of \(T\)-invariant functions given in Proposition 4.1 for step skew products over Markov shifts in case the latter arises from a strictly irreducible Markov kernel. It will turn out that in this situation the \(T\)-invariant functions only depend on the \(X\)-argument.

4.2 Theorem. Let \((\Omega, \nu, S)\) be a Markov shift with state space \(E\) and \((T_y)_{y \in E}\) be a measurable family of measure preserving transformations on a probability space \((X, \mu)\). Let \(T\) denote the respective step skew product. Assume that the Markov measure \(\nu\) arises from a Markov kernel \(\pi\) and a \(\pi\)-invariant probability measure \(m\) such that \(\pi\) is strictly irreducible with respect to \(m\). Then for every \(T\)-invariant function \(\varphi \in L^1(\nu \otimes \mu)\) there is a function \(f \in L^1(\mu)\) such that \(\varphi = \mathbb{1} \otimes f\).

Proof: It will suffice to prove the above statement for all characteristic functions of \(T\)-invariant sets. This is due to the fact that a function is \(T\)-invariant if and only if it is \(\Sigma_T\)-measurable, so every \(T\)-invariant function is the pointwise limit of linear combinations of such characteristic functions.

Let \(\varphi := \mathbb{1}_D\) be the characteristic function of a \(T\)-invariant set \(D\) and set \(\psi := \mathbb{1}_{D_x}\). Then Proposition 4.1 provides us with functions \(\hat{\varphi}, \hat{\psi} \in L^1(\nu \otimes \mu)\) such that for \(\nu \otimes \mu\)-almost all \((\omega, x) \in \Omega \times X\) we have

\[\varphi(\omega, x) = \hat{\varphi}(\omega_0, x)\]

and

\[\psi(\omega, x) = \hat{\psi}(\omega_0, x)\]

This implies that there is a set \(B \in \mathcal{E} \otimes \mathcal{A}\) such that \(\hat{\varphi} = \mathbb{1}_B\) and \(\hat{\psi} = \mathbb{1}_{B_x}\). It remains to show that up to a \(\nu \otimes \mu\)-null set we have \(B = E \times A\) for some set \(A \in \mathcal{A}\), which gives the claimed representation of \(\varphi\) with \(f := \mathbb{1}_A\).

To this end observe that by (4.3) a property holds \(\nu\)-almost surely if and only if it holds \(\nu_y\)-almost surely for \(m\)-almost all \(y \in E\). Therefore the \(T\)-invariance of \(\varphi\) implies that for \(m \otimes \mu\)-almost all \((y, x) \in E \times X\) we have

\[\varphi(\omega, x) = \varphi(S\omega, T_{x_0}x)\]

\(\nu_y\)-almost surely. Setting \(B_x := \{y \in E: (y, x) \in B\}\) we obtain thus

\[
\hat{\varphi}(y, x) = \mathbb{1}_{B_x}(y) = \nu_y(B_x \times \Omega) = \int_\Omega \mathbb{1}_{B_x}(\omega_0) \, d\nu_y(\omega) = \int_\Omega \hat{\varphi}(\omega_0, x) \, d\nu_y(\omega)
\]

\[
= \int_\Omega \varphi(\omega, x) \, d\nu_y(\omega) = \int_\Omega \varphi(S\omega, T_{x_0}x) \, d\nu_y(\omega)
\]

\[
= \int_{E_x} \hat{\varphi}(\omega_1, T_{y_0}x) \, d\nu_y(\omega) = \int_{E_x} \hat{\varphi}(\omega, T_{x_0}x) \, d\nu_y(\omega)
\]

for \(m \otimes \mu\)-almost all \((y, x) \in E \times X\) by 4.1 and 4.2. A similar argument shows that

\[
\hat{\psi}(y, x) = \int_{E_x} \hat{\psi}(\omega, T_{x_0}x) \, d\nu_y(\omega)
\]
Proof: (i) $\Rightarrow$ (ii): Observing that the measurable rectangles in $\Sigma_S \otimes \Sigma$ are $T$-invariant we immediately obtain that $\Sigma_S \otimes \Sigma \subseteq \Sigma_T$. So it remains to show that up to $\nu \otimes \mu$-null sets we have $\Sigma_T \subseteq \Sigma_S \otimes \Sigma$. To this end fix a set $C \in \Sigma_T$ and consider the characteristic...
Ergodicity of skew products

function $\varphi := 1_C$. Then by Theorem 4.12 we have $\varphi = 1 \otimes f$ for some $f \in L^1(\mu)$. Obviously $f = 1_A$ for some set $A \in \mathcal{F}$ and thus $C = \Omega \times A$ up to a $\nu \otimes \mu$-null set. We have to show that $f$ and thus $A$ is invariant under the family $(T_y)_{y \in E}$. To this end we compute

$$
\int_E \int_X |f \circ T_y(x) - f(x)| \, d\mu(x) \, d\nu(y) = \int_\Omega \int_X |f \circ T_{\omega y}(x) - f(x)| \, d\mu(x) \, d\nu(\omega) = \int_{\Omega \times X} |\varphi \circ T - \varphi| \, d\nu \otimes \mu = 0.
$$

This implies $f \circ T_y = f$ for $m$-almost all $y \in E$, which shows the claim.

(ii) $\Rightarrow$ (iii) Assume that $S$ is not ergodic. Then by [13] Theorem 5.2.11 there is an absorbing set $B \in \mathcal{F}$ with $m(B) \in (0, 1)$. Set $C := B \times \Omega$. Then $\nu(C) = m(B) \in (0, 1)$. Moreover, we have

$$
\nu(C \cap S^{-1}C^c) = \nu(B \times B^c \times \Omega) = \int_B \pi(y, B^c) \, dm(y) = 0.
$$

Since $S$ is $\nu$-preserving this shows that $C$ is $S$-invariant. Now consider the 2-point set $\mathcal{X} := \{1, 2\}$ equipped with the equidistributed probability measure $\bar{\nu}$ and let $\sigma$ be the non-trivial permutation of $\mathcal{X}$. We may define a measurable family $(T_y)_{y \in E}$ of measure preserving transformations by setting $T_y := \text{Id}$ if $y \in B$ and $T_y := \sigma$ if $y \in B^c$. It is not difficult to see that $\Sigma = \{\emptyset, \mathcal{X}\}$, so we get $\Sigma_S \otimes \Sigma = \{U \times \mathcal{X} : U \in \Sigma_S\}$. In particular the set $D := C \times \{1\}$ differs from any set in $\Sigma_S \otimes \Sigma$ by a set of positive measure. However, using the $S$-invariance of $C$, it is easy to see that $D$ is $T$-invariant and thus in $\Sigma_T$. This contradicts (ii). Therefore $S$ has to be ergodic. If in addition the family $(T_y)_{y \in E}$ is ergodic, then $\Sigma_S$ and $\Sigma$ are trivial. But by (ii) this implies that $\Sigma_T$ is trivial, so $T$ is ergodic.

(iii) $\Rightarrow$ (i) Assume that $\pi$ is not strictly irreducible with respect to $m$. If $\pi$ is also not irreducible with respect to $m$, then $S$ is not ergodic and we are done. So we may assume that $\pi$ is irreducible with respect $m$ and thus that $S$ is ergodic. We shall construct an ergodic measurable family $(T_y)_{y \in E}$ of measure preserving transformations such that the corresponding skew product $T$ is not ergodic.

To this end consider again the 2-point set $\mathcal{X} := \{1, 2\}$ equipped with the equidistributed $\bar{\nu}$ and let $\sigma$ denote the non-trivial permutation of $\mathcal{X}$. Then $\sigma$ is measure preserving and ergodic with respect to $\bar{\nu}$. Since $\pi$ is not strictly irreducible with respect to $m$, there is a deterministic set $B \in \mathcal{F}$ with probability $m(B) \in (0, 1)$. Consider the sets $B_1, \ldots, B_4$ given by

$$
B_1 := \{y \in B : \pi(y, B) = 1\}, \quad B_2 := \{y \in B : \pi(y, B) = 0\}
$$

and

$$
B_3 := \{y \in B^c : \pi(y, B^c) = 1\}, \quad B_4 := \{y \in B^c : \pi(y, B^c) = 0\}.
$$

Since the maps $\pi(\cdot, B)$ and $\pi(\cdot, B^c)$ are measurable by assumption, the sets $B_1, \ldots, B_4$ build a disjoint measurable partition of $E$. Based on this we may define a family $(T_y)_{y \in E}$ of $\bar{\nu}$-preserving transformations $T_y$ on $\mathcal{X}$ by setting $T_y := \sigma$ if $y \in B_2 \cup B_4$ and $T_y := \text{Id}$ if $y \in B_1 \cup B_3$. By the ergodicity of $S$ either $B_2$ or $B_4$ must have positive measure (it is easily seen that otherwise $B \times \Omega$ would be a non-trivial $S$-invariant set). This implies
that every measurable set $A \subseteq X$, which is invariant under the family $(T_y)_{y \in E}$, is also
invariant under $\sigma$ and thus trivial. In particular, the family is ergodic.

Let $T$ denote the step skew product arising from $S$ and $(T_y)_{y \in E}$. We claim that $T$ is
not ergodic. To show this we consider the sets $D_1, \ldots, D_4 \subseteq \Omega \times X$ given by

$$D_1 := (B_1 \times \Omega) \times \{1\}, \quad D_2 := (B_2 \times \Omega) \times \{1\},$$

$$D_3 := (B_3 \times \Omega) \times \{2\}, \quad D_4 := (B_4 \times \Omega) \times \{2\}$$

and set $D := D_1 \cup \cdots \cup D_4$. By the definition of the sets $B_1, \ldots, B_4$ we obtain

$$D = \{(B \times \Omega) \times \{1\} \cup ((B^c \times \Omega) \times \{2\}).$$

We claim that $D$ is a $T$-invariant, non trivial set. To see this observe that

$$\nu \otimes \pi(D) = m(B)\pi(1) + m(B^c)\pi(2) = \frac{1}{2},$$

So it remains to show that $D$ is $T$-invariant. Let $D^*_1$ denote the set $(B_1 \times B \times \Omega) \times \{1\}$. It is easy to see that $TD^*_1 = (B \times \Omega) \times \{1\} \subseteq D$. Moreover, we have

$$\nu \otimes \pi(D^*_1) = \nu(B_1 \times B \times \Omega) \pi(1) = \frac{1}{2} \int_{B_1} \pi(y, B) \ dm(y)$$

$$= \frac{1}{2} m(B_1) = \nu(B_1 \times \Omega) \pi(1) = \nu \otimes \pi(D_1).$$

Since $D^*_1 \subseteq D_1$ this shows that $\nu \otimes \pi(D_1 \cap T^{-1}D^c) = 0$. By a similar argument we obtain $\nu \otimes \pi(D_3 \cap T^{-1}D^c) = 0$. Let $D^*_2$ denote the set $(B_2 \times B^c \times \Omega) \times \{1\}$. Then we have $TD^*_2 = (B^c \times \Omega) \times \{2\} \subseteq D$. Moreover, we obtain

$$\nu \otimes \pi(D^*_2) = \nu(B_2 \times B^c \times \Omega) \pi(1) = \frac{1}{2} \int_{B_2} \pi(y, B^c) \ dm(y)$$

$$= \frac{1}{2} m(B_2) = \nu(B_2 \times \Omega) \pi(1) = \nu \otimes \pi(D_2).$$

Since $D^*_2 \subseteq D_2$ this shows that $\nu \otimes \pi(D_2 \cap T^{-1}D^c) = 0$. By a similar argument we obtain $\nu \otimes \pi(D_4 \cap T^{-1}D^c) = 0$. Noting that

$$D \cap T^{-1}D^c = \bigcup_{i=1}^4 (D_i \cap T^{-1}D^c)$$

we may conclude that $\nu \otimes \pi(D \cap T^{-1}D^c) = 0$. Since $T$ is measure preserving, this implies
that $D$ is $T$-invariant.

\section{Random ergodic theorems}

The study of random ergodic theorems has a long history. The first random ergodic
theorems were proposed by Pitt in \cite{Pitt} and by Ulam and von Neumann in \cite{Ulam}. Here,
a finite set of transformations is considered, which are chosen independently according
to some fixed probability distribution. Kakutani extended these results to the case of
arbitrarily many invertible transformations but still under the assumption of an i.i.d.
Ergodicity of skew products

selection process, cf. [20] Theorem 1]. Moreover he could identify the limit as the conditional expectation under the \( \sigma \)-algebra of sets, which are invariant under the family of transformations. The assumption of invertibility was removed by Ryll-Nardzewski in [31]. Kin proved convergence of random averages in general random dynamical systems, see [23]. However, there is no explicit representation of the limit function. Finally, the results of Bufetov in [6] give rise to a random ergodic theorem for finitely many transformations, which are chosen according to a stationary ergodic Markov chain. The limit may be identified as the above mentioned conditional expectation in case of a strictly irreducible transition matrix. The results obtained in the last section allow us to extend this result to random transformations driven by arbitrary stationary Markov chains. The random ergodic theorem of Kakutani and Ryll-Nardzewski is obtained as a special case by considering Bernoulli shifts and noting that trivial Markov kernels are always strictly irreducible with respect to their invariant probability measure.

5.1 Theorem. Consider a Markov kernel \( \pi \) and a \( \pi \)-invariant probability measure \( m \) on \( E \). Let \( (\Omega, \nu, S) \) denote the arising Markov shift. Consider furthermore a measurable family \( (T_y)_{y \in E} \) of measure preserving transformations on a probability space \( (X, \mu) \). Then for every function \( f \in L^1(\mu) \) there is a function \( \hat{f} \in L^1(m \otimes \mu) \) such that for \( \nu \)-almost all \( \omega \in \Omega \) the ergodic averages

\[
A_n f(\omega, \cdot) = \frac{1}{n} \sum_{i=0}^{n-1} f \circ T_i(\omega, \cdot)
\]

of \( f \) along the random iterations \( T_i := T_{\omega,x} \circ \cdots \circ T_{\omega,y} \) converge to \( \hat{f}(\omega_0, \cdot) \) in \( L^1 \) and \( \mu \)-almost surely. Moreover, if \( \pi \) is strictly irreducible with respect to \( m \), then for \( m \)-almost all \( y \in E \) the function \( \hat{f}(y, \cdot) \) coincides with the conditional expectation of \( f \) under \( \Sigma \).

Proof: For a given function \( f \in L^1(\mu) \) we shall consider the functions \( f^* := 1 \otimes f \) and \( \bar{f} := \mathbb{E}[f^* | \Sigma_T] \), which are obviously in \( L^1(\nu \otimes \mu) \). Then \( \bar{f} \) is \( T \)-invariant, so by Proposition[14] there is a function \( \hat{f} \in L^1(m \otimes \mu) \) such that

\[
\hat{f}(\omega, x) = \bar{f}(\omega_0, x)
\]

for \( \nu \otimes \mu \)-almost all \( (\omega, x) \in \Omega \times X \). Writing

\[
A_n f(\omega, x) = \frac{1}{n} \sum_{i=0}^{n-1} f^* \circ T_i(\omega, x)
\]

we obtain thus

\[
\lim_{n \to \infty} A_n f(\omega, x) = \bar{f}(\omega, x) = \hat{f}(\omega_0, x)
\]

for \( \nu \otimes \mu \)-almost all \( (\omega, x) \in \Omega \times X \) by Birkhoff’s ergodic theorem.

For the mean convergence we shall first assume that \( f \in L^2(\mu) \). Then obviously \( f^* \in L^2(\mu) \) and \( \| f^* \|_2 \leq \max\{1, \| f \|_2\} \). It is also not difficult to see that \( \bar{f} \in L^2(\nu \otimes \mu) \) with \( \| \bar{f} \|_2 \leq \| f^* \|_2 \). Using these observations and the \( T \)-invariance of \( \bar{f} \) we obtain

\[
| A_n f(\omega, x) - \hat{f}(\omega_0, x) | = \left| \frac{1}{n} \sum_{i=0}^{n-1} f^* \circ T_i(\omega, x) - \frac{1}{n} \sum_{i=0}^{n-1} \bar{f} \circ T_i(\omega, x) \right|
\]

\[
\leq \frac{1}{n} \sum_{i=0}^{n-1} |(f^* - \bar{f}) \circ T_i(\omega, x)| \leq \| f^* - \bar{f} \|_2
\]

\[
\leq 2 \| f^* \|_2 \leq 2 \max\{1, \| f \|_2\}
\]

17
for \( \nu \otimes \mu \)-almost all \((\omega, x) \in \Omega \times X\). Since the left hand side converges \( \mu \)-almost surely to zero for \( \nu \)-almost every \( \omega \in \Omega \), we may apply dominated convergence to obtain
\[
\lim_{n \to \infty} \| A_n f(\omega, \cdot) - \hat{f}(\omega_0, \cdot) \|_1 = 0
\]
for every such \( \omega \in \Omega \). Now for a general \( f \in L^1(\mu) \) a standard 3ε-argument shows that for \( \nu \)-almost all \( \omega \in \Omega \) the averages \( A_n f(\omega, \cdot) \) form a Cauchy sequence and converge therefore to some limit in \( L^1(\mu) \), which equals \( \hat{f}(\omega_0, \cdot) \) by \( 5.2 \).

Finally, if \( \pi \) is strictly irreducible with respect to \( m \), then \( \pi \) is in particular irreducible with respect to \( m \) by Proposition \([5,3]\). Accordingly, by Proposition \([5,2]\) the shift \( S \) is ergodic, thus \( \Sigma_S \) is trivial. Using this we may apply statement (ii) of Theorem \([4,3]\) to obtain that
\[
\hat{f}(\omega_0, x) = \mathbb{E}[f^*|\Sigma_T](\omega, x) = \mathbb{E}[(1 \otimes f)|\Sigma_S \otimes \Sigma](\omega, x) = \mathbb{E}[(1 \otimes f)|\Sigma](\omega, x)
\]
for \( \nu \otimes \mu \)-almost all \((\omega, x) \in \Omega \times X\), which shows the claim by the fact that \( m \) is the push-forward measure of \( \nu \) under the projection \( \xi_0 \) to the coordinate 0.

We may also give an integrated version of the above random ergodic theorem. For finite state Markov chains such a theorem was proved by Bufetov, cf. \([6, \text{Theorem 3}]\) and \([3, \text{Corollary 2}]\). More precisely, instead of the time averaging operators \( A_n \) we shall consider the expectation operators \( M_n \) given by
\[
M_n f(x) := \int_{\Omega} f \circ T^n_x \ d\nu(\omega)
\]
for \( n \in \mathbb{N}_0, f \in L^1(\mu) \) and \( x \in X \). In case of a finite state space the operator \( M_n \) describes just the weighted sum of \( f \) over all possible compositions of transformations of length \( n \). Recall the definition of the space \( L \log L(\mu) \), which consists of all measurable functions \( f \) such that \( \int |f| \log^+ |f| \ d\mu < \infty \).

5.2 Theorem. Consider a Markov kernel \( \pi \) and a \( \pi \)-invariant probability measure \( m \) on \( E \). Let \((\Omega, \nu, S)\) denote the arising Markov shift. Consider furthermore a measurable family \((T_y)_{y \in E}\) of measure preserving transformations on a probability space \((X, \mu)\). Then for every function \( f \in L^1(\mu) \) there is a function \( \tilde{f} \in L^1(\mu) \) such that we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} M_j f = \tilde{f}
\]
in the \( L^1 \)-sense. If \( f \in L \log L(\mu) \), we obtain in addition \( \mu \)-almost sure convergence. Finally, if \( \pi \) is strictly irreducible with respect to \( m \), then the limit \( \tilde{f} \) equals the conditional expectation of \( f \) under \( \Sigma \).

Proof: Fix \( f \in L^1(\mu) \). Then by Theorem \([5,4]\) we have \( \nu \otimes \mu \)-almost sure convergence of the averages \( A_n f \) to a limit \( \tilde{f} \) in \( L^1(\nu \otimes \mu) \). Setting \( f^* := 1 \otimes f \) and \( \bar{f} := \mathbb{E}[f^*|\Sigma_T] \) we obtain as above \( \bar{f}(\omega_0, x) = \bar{f}(\omega_0, x) \) for \( \nu \otimes \mu \)-almost all \((\omega, x) \in \Omega \times X \). Note that by Fubini’s theorem the partial integral
\[
\tilde{f}(\cdot) := \int_E \bar{f}(y, \cdot) \ dm(y) = \int_{\Omega} \bar{f}(\omega, \cdot) \ d\nu(\omega)
\]
defines a function in $L^1(\mu)$. (The last equality follows again from the fact that $m$ is the push-forward of $\nu$ under the projection $\xi_0$.) Thus we obtain

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} M_j f - \tilde{f} \right\|_1 = \int_X \left( \int_\Omega \left( \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j_\omega(x) - \tilde{f}(\omega, x) \right) \, d\nu(\omega) \right) \, d\mu(x)$$

$$\leq \int_\Omega \int_X \left( \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j_\omega(x) - \tilde{f}(\omega, x) \right) \, d\mu(x) \, d\nu(\omega)$$

$$= \int_\Omega \left( \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j_\omega(x) - \tilde{f}(\omega, x) \right) \, d\nu \otimes \mu(\omega, x)$$

$$= \left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - \tilde{f} \right\|_1,$$

where the latter term converges to zero by the mean ergodic theorem. If $\pi$ is strictly irreducible with respect to $m$, we have furthermore

$$\tilde{f}(x) = \int_E \tilde{f}(y, x) \, d\mu(y) = \int_E \mathbb{E}[f|\Sigma](x) \, d\mu(y) = \mathbb{E}[f|\Sigma](x)$$

for $\mu$-almost all $x \in X$ by Theorem 5.1

Now assume $f \in L \log L(\mu)$. Then $f$ is integrable with respect to $\mu$, which in turn gives that $f^*$ and therefore $A_n f$ is integrable with respect to $\nu \otimes \mu$ for all $n \in \mathbb{N}$. Again by Fubini’s theorem we may conclude that $A_n f(x)$ is a function in $L^1(\nu)$ satisfying

$$\int_\Omega A_n f(\omega, x) \, d\nu(\omega) = \frac{1}{n} \sum_{j=0}^{n-1} \int_\Omega f \circ T^j_\omega(x) \, d\nu(\omega) = \frac{1}{n} \sum_{j=0}^{n-1} M_j f(x)$$

(5.3)

for $\mu$-almost all $x \in X$. Consider the maximal function $F$ given by

$$F(\omega, x) := \sup_{n \in \mathbb{N}} |A_n f(\omega, x)|$$

for $(\omega, x) \in \Omega \times X$. Since $f \in L \log L(\mu)$, we have $f^* \in L \log L(\nu \otimes \mu)$. Therefore, by [5.1] and Wiener’s dominated ergodic theorem, cf. [25] Chapter 1, Theorem 6.3, we obtain $F \in L^1(\nu \otimes \mu)$. This implies that $F(\cdot, x)$ is in $L^1(\nu)$ for $\mu$-almost all $x \in X$. In particular, for $\mu$-almost all $x \in X$ the functions $A_n f(\cdot, x)$ converge $\nu$-almost surely to $\tilde{f}(\cdot, x)$ while being dominated by the integrable function $F(\cdot, x)$. Dominated convergence yields thus

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} M_j f(x) = \int_\Omega \tilde{f}(\omega, x) \, d\nu(\omega) = \tilde{f}(x)$$

for $\mu$-almost all $x \in X$ by [5.3].

**Outlook.** It is well known that in the case of an i.i.d. selection of transformations the assumptions of the random ergodic theorem may be weakened. In fact, instead of a common invariant probability measure it suffices to consider a probability measure, which is stationary for the family of transformations, cf. [22] Chapter 1.2. The latter means that the measure is invariant under the Markov kernel induced by the random
transformations on the state space of the RDS. In fact, one can show that a measure is stationary if and only if the respective product measure is invariant under the skew product. This observation is particular useful in situations where the transformations are not invertible and the assumption of a common invariant measure excludes interesting behavior (as it is often the case in low dimensional dynamics). However, if we allow the selection process to be Markov chain, a similarly convenient correspondence is not at hand, so it seems reasonable to confine to invariant probability measures. As discussed in the introduction in many applications this assumption is naturally instantiated. However, in certain contexts it is more or less restrictive, see e.g. [26]. It would therefore be interesting, to what extend this assumption can be relaxed also in the present setting.

In the case of a finite state Markov chain Theorem 5.2 can be improved in the sense that pointwise convergence holds for all $L^1$-functions, cf. [8, Corollary 2]. This is due to the fact that the averages over all paths of the Markov chain with a fixed first symbol can be encoded by a convenient Markov operator (which is also implicitly present in the computations of the proof of Theorem 4.2). Since in this case the averages considered in Theorem 5.2 are just a finite weighted sum of the ergodic averages of this operator, one may use the general ergodic theorem for Dunford-Schwartz operators to obtain pointwise convergence for all $L^1$-functions. In the general setting this approach seems to fail. A priori there is no argument at hand allowing to interchange integrating and taking the limit of the operator ergodic averages. It would therefore be interesting, if there is an alternative argument showing pointwise convergence for $L^1$-functions.

Acknowledgements. The authors would like to thank Alexander I. Bufetov and Anush Tserunyan for stimulating discussions, valuable hints and helpful remarks. Felix Pogorzelski und Elias Zimmermann gratefully acknowledge support by the German Israeli Foundation for Scientific Research and Development (GIF, grant I-1485-304.6/2019).

References

[1] J. Aaronson: An introduction to infinite ergodic theory, Mathematical Surveys and Monographs 50, American Mathematical Society, 1997
[2] J. Aaronson and M. Lemańczyk: Exactness of Rokhlin endomorphisms and weak mixing of Poisson boundaries, in: Algebraic and topological dynamics, Contemporary Mathematics 385, American Mathematical Society, 77-87, 2005
[3] L. Arnold: Random dynamical systems, Springer Monographs in Mathematics, Springer, 1998
[4] L. Bowen and A. Nevo: Amenable equivalence relations and the construction of ergodic averages for group actions, J. Anal. Math. 126, 359-388, 2015
[5] Z. I. Bezhaeva and V. I. Oseledets: On a symmetric $\sigma$-algebra of a stationary Markov Harris process, Theory Probab. Appl. 41(4), 741-748, 1996
[6] A. I. Bufetov: Skew products and ergodic theorems for group actions, Zap. Nauchn. Semin. POMI 266(5), 13-28, 2000. Translated in: J. Math. Sci. 113(4), 548-557, 2003
[7] A. I. Bufetov: Operator ergodic theorems for actions of free semigroups and groups, Funct. Anal. Appl. 34(4), 239-251, 2000
[8] A. I. Bufetov: Markov averaging and ergodic theorems for several operators, in: Topology, ergodic theory, real algebraic geometry, Translations Series 2, American Mathematical Society, 39-50, 2001
Ergodicity of skew products

[9] A. I. Bufetov and A. V. Klimenko: On Markov operators and ergodic theorems for group actions, Eur. J. Comb. 33(7), 1427-1443, 2012
[10] A. I. Bufetov and A. V. Klimenko: Maximal inequality and ergodic theorems for Markov groups, Proc. Steklov Inst. Math., 277, 27-42, 2012
[11] A. I. Bufetov, M. I. Khristoforov and A. V. Klimenko: Cesàro convergence of spherical averages for measure-preserving actions of Markov semigroups and groups, Int. Math. Res. Not. 21, 4797-4829, 2012.
[12] A. I. Bufetov and C. Series: A pointwise ergodic theorem for Fuchsian groups, Math. Proc. Camb. Philos. Soc. 151(1), 145-159, 2011
[13] R. Douc, E. Moulines, P. Priouret and P. Soulier: Markov chains, Springer Series in Operations Research and Financial Engineering, Springer, 2018.
[14] T. Eisner, B. Farkas, M. Haase and R. Nagel: Operator theoretic aspects of ergodic theory, Graduate Texts in Mathematics 272, Springer, 2015
[15] A. Furman: Random walks on groups and random transformations in: Handbook of dynamical systems 1A, North Holland, 931-1014, 2002
[16] K. Fujiwara and A. Nevo: Maximal and pointwise ergodic theorems for word-hyperbolic groups, Ergodic Theory Dyn. Syst. 18(4), 843-858, 1998
[17] R. I. Grigorchuk: Individual ergodic theorems for actions of free groups (Russian), Proceedings of the Tambov Workshop in the Theory of Functions, 1986
[18] R. I. Grigorchuk: Ergodic theorems for actions of free groups and free semigroups, Math. Notes 65(5), 654-657, 1999
[19] L. A. Grigorenko: On the σ-algebra of symmetric events for a countable Markov chain, Theory Probab. Appl. 24, 199-204, 1979
[20] S. Kakutani: Random ergodic theorems and Markoff processes with a stable distribution, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability 1950, University of California Press, 247-261, 1951
[21] A. S. Kechris: Classical descriptive set theory, Graduate Texts in Mathematics 156, Springer, 1995
[22] Y. Kifer: Ergodic theory of random transformations, Progress in Probability and Statistics 10, Birkhäuser, 1986
[23] E. Kin: The general random ergodic theorem, Z. Wahrscheinlichkeitstheor. Verw. Geb. 22, 120-135, 1972
[24] Z. S. Kowalski: Frobenius-Perron operator description of Markov chains, Prob. Math. Stat. 35(2), 325-341, 2015
[25] U. Krengel: Ergodic theorems, de Gruyter Studies in Mathematics 6, de Gruyter, 1985
[26] E. Matias: Markovian random iterations of homeomorphisms of the circle, Ergodic Theory Dyn. Syst. 42(9), 2953-2956, 2022
[27] A. Nevo, E. M. Stein: A generalization of Birkhoff’s pointwise ergodic theorem, Acta Math. 173(1), 135-154, 1994
[28] E. Nummelin: General irreducible Markov chains and non-negative operators, Cambridge Tracts in Mathematics 83, Cambridge University Press, 1984
[29] V. I. Oseledets: Markov chains, skew products and ergodic theorems for “general” dynamic systems, Theory Probab. Appl. 10, 499-504, 1965
[30] H. R. Pitt: Some generalizations of the ergodic theorem, Proc. Camb. Philos. Soc. 38, 325-343, 1942

21
[31] C. Ryll-Nardzewski: *On the ergodic theorems. III: The random ergodic theorems*, Stud. Math. 14, 298–301, 1955

[32] A. Tserunyan, J. Zomback: *Characterization of weak mixing for boundary actions of free semigroups*, manuscript in preparation

[33] S. M. Ulam and J. von Neumann: *Random ergodic theorems*, Bull. Amer. Math. Soc. 51, 660, 1945

Mathematical Institute, University of Leipzig
Augustusplatz 10, 04109 Leipzig
pablo.lummerzheim@protonmail.com
felix.pogorzelski@math.uni-leipzig.de
elias.zimmermann@math.uni-leipzig.de