Convergence of Markov chain transition probabilities

Michael Scheutzow*   Dominik Schindler†

Abstract

Consider a discrete time Markov chain with rather general state space which has an invariant probability measure μ. There are several sufficient conditions in the literature which guarantee convergence of all or μ-almost all transition probabilities to μ in the total variation (TV) metric: irreducibility plus aperiodicity, equivalence properties of transition probabilities, or coupling properties. In this work, we review and improve some of these criteria in such a way that they become necessary and sufficient for TV convergence of all respectively μ-almost all transition probabilities. In addition, we discuss so-called generalized couplings.

1 Introduction

It is a classical result that all transition probabilities of a discrete time Markov chain with invariant probability measure (ipm) μ on a rather general state space E converge to μ in the total variation metric provided that the chain is irreducible, recurrent and aperiodic ([12]). Further, *Doob's theorem* states that under appropriate additional conditions, ultimate equivalence of every pair of transition probabilities implies the same result (see [3, Theorem 4.2.1] or [10]). Finally the existence of *couplings* of chains starting at different initial conditions entails total variation convergence to μ. The goal of this paper is to modify the sufficient conditions in the literature in such a way that they become equivalent. It will turn out, for example, that *asymptotic equivalence* of transition probabilities (which seems to be a new concept) is equivalent to total variation convergence of all transition probabilities. It is also of interest to find weaker conditions which only imply total variation convergence of the transition probabilities starting from μ-almost every $x \in E$. Again we will provide necessary and sufficient conditions similar to those described above. We will also address a convergence property strictly between these two and again we will provide necessary and sufficient conditions. Apart from couplings we will also formulate equivalent conditions in terms of *generalized couplings* for each of the convergence properties.

---

*Institut für Mathematik, MA 7-5, Fakultät II, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany. E-mail: ms@math.tu-berlin.de

†Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom. E-mail: dominik.schindler19@imperial.ac.uk

Throughout this paper $(E, \mathcal{E})$ denotes a measurable space for which $\mathcal{E}$ is countably generated and the diagonal $\Delta := \{(x, x) : x \in E\}$ is in $\mathcal{E} \otimes \mathcal{E}$ (or, equivalently, $\mathcal{E}$ is countably generated and separates points or, equivalently, $\mathcal{E}$ is countably generated and all singletons are in $\mathcal{E}$ (see [5, p. 116])). Let P be a Markov kernel on E and denote the corresponding n-step transition probability by $P_n(.,.)$, $n \in \mathbb{N}_0$. $\mathbb{P}_x$ denotes the law of the Markov chain starting at $x \in E$. Note that $\mathbb{P}_x$ is a probability measure on $(E^{\mathbb{N}_0}, \mathcal{E}^{\otimes \mathbb{N}_0})$. We will often identify a Markov chain and its Markov kernel P and denote the corresponding Markov chain by X. We denote the *total variation* metric on the space of probability measures on $(E, \mathcal{E})$ by d, i.e. $d(\nu_1, \nu_2) := \sup_{A \in \mathcal{E}} |\nu_1(A) - \nu_2(A)|$. We say that $P_n(x, .)$ *converges* to a probability measure μ on $(E, \mathcal{E})$ if $P_n(x, .)$ converges to μ in the total variation metric as $n \to \infty$. Throughout the paper we will assume that P admits an ipm μ (but we will not assume uniqueness of μ). From now on, the letter μ will always denote an invariant probability measure of the Markov chain X associated to P.

Let ν_1 and ν_2 be measures on the same measurable space $(\bar{E}, \bar{\mathcal{E}})$. Then we say (as usual) that ν_1 is absolutely continuous with respect to ν_2 (notation $\nu_1 \ll \nu_2$) if $A \in \bar{\mathcal{E}}$ with $\nu_2(A) = 0$ implies $\nu_1(A) = 0$, and that ν_1 and ν_2 are equivalent (denoted $\nu_1 \sim \nu_2$) if they are mutually absolutely continuous. Further we write $\nu_1 \not\perp \nu_2$ if ν_1 and ν_2 are non-singular, i.e. there does not exist a set $A \in \bar{\mathcal{E}}$ such that $\nu_1(A) = 0$ and $\nu_2(A^c) = 0$, where A^c denotes the complement of the set A. Any measure ξ on $(\bar{E} \times \bar{E}, \bar{\mathcal{E}} \otimes \bar{\mathcal{E}})$ with marginals ν_1 and ν_2 is called a *coupling* of ν_1 and ν_2. We write $\xi \in C(\nu_1, \nu_2)$. Recall the *coupling equality*: for probability measures ν_1 and ν_2 on $(\bar{E}, \bar{\mathcal{E}})$, we have $d(\nu_1, \nu_2) = \inf\{\xi(\Delta^c) : \xi \in C(\nu_1, \nu_2)\}$ ([9, Theorem 2.2.2]). We will call a pair (X, Y) of $\bar{E}$-valued random variables defined on the same probability space a coupling of the probability measures ν_1 and ν_2 on $(\bar{E}, \bar{\mathcal{E}})$, if their joint law is a coupling of ν_1 and ν_2. Below we will deal with the cases $\bar{E} := E$ and $\bar{E} := E^{\mathbb{N}_0}$. We will define the concept of a *generalized coupling* later. Generalized (asymptotic) couplings are particularly useful to prove *weak* convergence of transition probabilities (see [11] and [2]) but (non-asymptotic) generalized couplings can also be used to establish upper bounds on the total variation distance of transition probabilities (see [6, Proof of Theorem 1.1]).

We will formulate all results in the discrete-time set-up. This is essentially without loss of generality. Indeed, assume that μ is an invariant probability measure of an E-valued continuous-time Markov process. Then μ is also an ipm of the associated *skeleton chain* sampled at times $0, h, 2h, ...$ and for each $x \in E$ total variation convergence of $P_{nh}(x, .)$ to μ (as $n \to \infty$) for some $h > 0$ is equivalent to total variation convergence of $P_t(x, .)$ to μ since $t \mapsto d(P_t(x, .), \mu)$ is non-increasing.

Once one has established convergence of all or almost all transition probabilities then it is natural to ask for the speed of convergence. A large number of papers have been devoted to these questions, for example [8], [14] and [9]. We will, however, not touch these questions here.

At some point we will need a stronger condition on the measurable space $(E, \mathcal{E})$: as usual, we say that $(E, \mathcal{E})$ is a *Borel* space if it is isomorphic (as a measurable space) to a Borel subset of $[0, 1]$. In particular, this holds for a complete, separable metric space E equipped with its Borel σ-field $\mathcal{E}$.

2 Necessary and sufficient conditions for total variation convergence

Let $(X_n)_{n \in \mathbb{N}_0}$ be a Markov chain with transition kernel P, ipm μ and state space $(E, \mathcal{E})$ as in the introduction. We adopt the following notation (cf. [12]).

Notation 2.1. *For $x \in E$, $A \in \mathcal{E}$,*

$$Q(x,A) := \mathbb{P}_x\big(\{X_n \in A \text{ for infinitely many } n \in \mathbb{N}\}\big),$$

$$L(x,A) := \mathbb{P}_x\Big(\bigcup_{n=1}^{\infty} \{X_n \in A\}\Big).$$

We start by defining three properties of increasing generality about the convergence of Markov chain transition probabilities which we will be interested in.

Properties 2.2. We say that

- Property P_1 holds if $P_n(x,.)$ converges to μ for every $x \in E$.
- Property P_2 holds if $P_n(x,.)$ converges to μ for μ-almost all $x \in E$ and $\lim_{n \to \infty} d(P_n(x,.),\mu) < 1$ for all $x \in E$.
- Property P_3 holds if $P_n(x,.)$ converges to μ for μ-almost all $x \in E$.

Remark 2.3. Note that Properties P_1 and P_2 both imply uniqueness of μ (we will show the latter claim in Remark 5.1). Note also that $\lim_{n \to \infty} d(P_n(x,.),\mu)$ always exists since μ is invariant and the total variation distance can never increase when applying a measurable map. Therefore, we could replace "$\lim_{n \to \infty} d(P_n(x,.),\mu) < 1$ for all $x \in E$" in P_2 by "for each x there exists some $n \in \mathbb{N}_0$ such that $d(P_n(x,.),\mu) < 1$" without changing the class of chains for which P_2 holds. One might also be interested in a modification $\tilde{P}_2$ of Property P_2 in which the last property $\lim_{n \to \infty} d(P_n(x,.),\mu) < 1$ for all $x \in E$ is replaced by uniqueness of μ. Clearly, P_2 is stronger than $\tilde{P}_2$ and it is easy to see that it is strictly stronger. Property $\tilde{P}_2$ was studied in [10], for example, but P_2 is more closely related to conditions studied in the literature. We will see, in particular, that the assumptions of [10, Corollary 1] do not only imply $\tilde{P}_2$ but even P_2. Example 5.2 shows that one cannot delete the first part of property P_2 without changing the class of chains for which it holds.

We will define four sets of assumptions, one in terms of equivalence or non-singularity of transition probabilities, one in terms of aperiodicity and recurrence or irreducibility properties, one in terms of couplings and one in terms of generalized couplings. It will turn out that all assumption with index i, $i \in \{1,2,3\}$, not only imply property P_i but are also *necessary* for P_i to hold. In some cases we formulate conditions with an additional prime (or some other symbol) which will formally be stronger than the same condition without prime but which will in fact turn out to be equivalent (at least when the state space is Borel). Before we state various assumptions we define the (possibly new) concept of asymptotic equivalence of transition probabilities.

Definition 2.4. *We say that the states $x \in E$ and $y \in E$ are asymptotically equivalent if for each $\varepsilon > 0$ there exists some $n \in \mathbb{N}$ and a set $A \in \mathcal{E}$ such that $P_n(x,A) \geq 1 - \varepsilon$, $P_n(y,A) \geq 1 - \varepsilon$, and the measures $P_n(x,.)$ and $P_n(y,.)$ restricted to the set A are equivalent.*

Remark 2.5. Note that if for given $x,y \in E$, $\varepsilon > 0$ and $n \in \mathbb{N}$ there exists a set A as in the previous definition, then there exists a set $\bar{A}$ as in the previous definition (with the same ε) if n is replaced by $n+1$ (and, by iteration, the same holds for all integers larger than n). This implies, in particular, that asymptotic equivalence induces an equivalence relation on E.

The first set of assumptions is formulated in terms of equivalence or non-singularity of transition probabilities.

Assumption 2.6. *We say that*

- *Assumption A_1 holds if all pairs $(x,y) \in E \times E$ are asymptotically equivalent.*

- *Assumption A_2 holds if for all $(x, y) \in E \times E$ there exists some $n = n_{x,y} \in \mathbb{N}$ such that $P_n(x, .) \not\perp P_n(y, .)$.*
- *Assumption A_3 holds if for $\mu \otimes \mu$-almost all $(x, y) \in E \times E$ there exists some $n = n_{x,y} \in \mathbb{N}$ such that $P_n(x, .) \not\perp P_n(y, .)$.*
- *Assumption A'_3 holds if $\mu \otimes \mu$-almost all $(x, y) \in E \times E$ are asymptotically equivalent.*

Lemma A.7 states that the set of all $(x, y) \in E \times E$ which are asymptotically equivalent is a measurable subset of $(E \times E, \mathcal{E} \otimes \mathcal{E})$.

Remark 2.7. Obviously, Property P_1 implies that any two states x, y are asymptotically equivalent (i.e. A_1 holds) while the simple Example 5.3 shows that it does *not* imply the stronger property "for all $x, y \in E$ there exists some $n = n_{x,y} \in \mathbb{N}_0$ such that $P_n(x, .) \sim P_n(y, .)$" under which P_1 was shown in [10, Theorem 1].

Before we state the second set of assumptions, we define the concepts of aperiodicity, irreducibility and the Harris property for a Markov kernel P with invariant measure μ.

Definition 2.8. *[14, p. 32] The Markov kernel P (with invariant probability measure μ) is called d-periodic, if $d \geq 2$, and there are disjoint sets $E_1, E_2, ..., E_d \in \mathcal{E}$ with $\mu(E_1) > 0$ that fulfill*

$$P(x, E_{i+1(\mathrm{mod}\, d)}) = 1 \quad \forall x \in E_i, 1 \leq i \leq d. \tag{2.1}$$

The chain is called aperiodic *if no such $d \geq 2$ exists.*

Definition 2.9. *The Markov kernel P is called ϕ-irreducible if ϕ is a non-trivial σ-finite measure on $(E, \mathcal{E})$ such that for all $A \in \mathcal{E}$ with $\phi(A) > 0$ and all $x \in E$ we have $L(x, A) > 0$ (or, equivalently, there exists some $n = n(x, A) \in \mathbb{N}$ such that $P_n(x, A) > 0$). P is called* irreducible *if P is ϕ-irreducible for some non-trivial ϕ. We say that P is* weakly irreducible *(with respect to the given ipm μ) if there exists some non-trivial σ-finite measure ϕ on $(E, \mathcal{E})$ and a set $E_0 \in \mathcal{E}$ satisfying $\mu(E_0) = 1$ such that for every $x \in E_0$ and every $A \in \mathcal{E}$ with $\phi(A) > 0$ we have $L(x, A) > 0$.*

Remark 2.10. It is straightforward to check that if ϕ is as in the definition (either part), then $\phi \ll \mu$. Further, if P is (weakly) μ-irreducible then P is (weakly) ϕ-irreducible for every non-trivial σ-finite measure on $(E, \mathcal{E})$ satisfying $\phi \ll \mu$. We will show in Proposition A.1 the less obvious fact that (ϕ-)irreducibility implies μ-irreducibility (which, in the terminology of [12, Proposition 4.2.2], means that μ is the *maximal irreducibility measure*). We will use Proposition A.1 only in the proof of Theorem 2.17.

Definition 2.11. *[12, p. 199] P or the associated Markov chain X are called* Harris *(or* Harris recurrent*), if there exists a non-trivial σ-finite measure ϕ on $(E, \mathcal{E})$ such that for all $A \in \mathcal{E}$ with $\phi(A) > 0$ and all $x \in E$ we have $Q(x, A) = 1$ (or, equivalently, $L(x, A) = 1$ for all $x \in E$ and $A \in \mathcal{E}$ with $\phi(A) > 0$).*

Assumption 2.12. *We say that*

- *Assumption B_1 holds if P is aperiodic and Harris.*
- *Assumption B_2 holds if P is aperiodic and irreducible.*
- *Assumption B_3 holds if P is aperiodic and weakly irreducible.*

Note that Harris recurrence implies irreducibility, so B_1 implies B_2.

Let $\mathcal{M}(\bar{E})$ be the set of all probability measures on the measurable space $(\bar{E}, \bar{\mathcal{E}})$. For $\xi \in \mathcal{M}(\bar{E} \times \bar{E})$, we denote the i-th marginal by ξ^i, $i \in \{1, 2\}$. If $(\bar{E}, \bar{\mathcal{E}}) = (E^{\mathbb{N}_0}, \mathcal{E}^{\mathbb{N}_0})$, then we denote the projection of ξ resp. ξ^i onto the k-th coordinate by ξ_k resp. ξ_k^i, $k \in \mathbb{N}_0$, $i \in \{1, 2\}$. We can now define the third set of assumptions in terms of couplings.

Assumption 2.13. *We say that*

- *Assumption C_1 holds if for each $x, y \in E$ and $m \in \mathbb{N}$ there exists some $k_m \in \mathbb{N}_0$ and a coupling $\zeta[m] \in C(P_{k_m}(x, .), P_{k_m}(y, .))$ such that $\zeta[m](\Delta) \geq 1 - \frac{1}{m}$.*

- *Assumption $\hat{C}_1$ holds if for each $x, y \in E$ and $m \in \mathbb{N}$ there exists a coupling $\zeta[m] \in C(P_m(x, .), P_m(y, .))$ such that $\lim_{m \to \infty} \zeta[m](\Delta) = 1$.*
- *Assumption $\mathring{C}_1$ holds if for each $x, y \in E$ there exists a coupling $\xi \in C(\mathbb{P}_x, \mathbb{P}_y)$ such that $\lim_{m \to \infty} \xi_m(\Delta) = 1$.*
- *Assumption C_1' holds if for each $x, y \in E$ there exists a coupling $(X_k)_{k \in \mathbb{N}_0}$, $(Y_k)_{k \in \mathbb{N}_0}$ of $\mathbb{P}_x$ and $\mathbb{P}_y$ on some space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\lim_{n \to \infty} \mathbb{P}(X_k = Y_k \text{ for all } k \geq n) = 1$.*
- *Assumption C_2 holds if for all $x, y \in E$ there exists some $k \in \mathbb{N}_0$ and a coupling $\zeta \in C(P_k(x, .), P_k(y, .))$ such that $\zeta(\Delta) > 0$.*
- *Assumption C_2' holds if for each $x, y \in E$ there exists a coupling $(X_k)_{k \in \mathbb{N}_0}$, $(Y_k)_{k \in \mathbb{N}_0}$ of $\mathbb{P}_x$ and $\mathbb{P}_y$ on some space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\liminf_{n \to \infty} \mathbb{P}(X_k = Y_k \text{ for all } k \geq n) > 0$ and for $\mu \otimes \mu$-almost every $(x, y) \in E \times E$ there exists a coupling $(X_k)_{k \in \mathbb{N}_0}$, $(Y_k)_{k \in \mathbb{N}_0}$ of $\mathbb{P}_x$ and $\mathbb{P}_y$ on some space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\lim_{n \to \infty} \mathbb{P}(X_k = Y_k \text{ for all } k \geq n) = 1$.*
- *Assumption C_3 holds if for $\mu \otimes \mu$-almost every $(x, y) \in E \times E$ there exists some $k \in \mathbb{N}_0$ and a coupling $\zeta \in C(P_k(x, .), P_k(y, .))$ such that $\zeta(\Delta) > 0$.*
- *Assumption C_3' holds if for $\mu \otimes \mu$-almost every $(x, y) \in E \times E$ there exists a coupling $(X_k)_{k \in \mathbb{N}_0}$, $(Y_k)_{k \in \mathbb{N}_0}$ of $\mathbb{P}_x$ and $\mathbb{P}_y$ on some space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\lim_{n \to \infty} \mathbb{P}(X_k = Y_k \text{ for all } k \geq n) = 1$.*

We chose Condition C_i such that it is as weak as possible and C_i' such that it is as strong as possible subject to the requirement that both are equivalent to all other conditions with the same index i (in case the state space is Borel). Note that there are several natural conditions in between C_i and C_i' ($i = 1, 2, 3$) for which there is no need to state them, since they will all turn out to be equivalent (at least in the Borel case).

Finally, we define the concept of a *generalized coupling* and formulate the fourth set of assumptions.

Definition 2.14. *For probability measures ν_1 and ν_2 on $(\bar{E}, \bar{\mathcal{E}})$, define*

- $\tilde{C}(\nu_1, \nu_2) := \{\xi \in \mathcal{M}(\bar{E} \times \bar{E}) : \xi^1 \ll \nu_1, \xi^2 \ll \nu_2\}$,
- $\check{C}(\nu_1, \nu_2) := \{\xi \in \mathcal{M}(\bar{E} \times \bar{E}) : \xi^1 \ll \nu_1, \xi^2 \sim \nu_2\}$.

Assumption 2.15. *We say that*

- *Assumption G_1 holds if for each pair $(x, y) \in E \times E$ there exists some $\xi \in \check{C}(\mathbb{P}_x, \mathbb{P}_y)$ such that $\lim_{k \to \infty} \xi_k(\Delta) = 1$,*
- *Assumption G_2 holds if for each pair $(x, y) \in E \times E$ there exists some $k \in \mathbb{N}$ and $\zeta \in \tilde{C}(P_k(x, .), P_k(y, .))$ such that $\zeta(\Delta) > 0$.*
- *Assumption G_3 holds if for $\mu \otimes \mu$-almost every $(x, y) \in E \times E$ there exists some $k \in \mathbb{N}$ and $\zeta \in \tilde{C}(P_k(x, .), P_k(y, .))$ such that $\zeta(\Delta) > 0$.*

Example 5.4 below shows that not only is it the case that G_2 does not imply G_1, but neither does the stronger version of G_2 in which > 0 is replaced by $= 1$.

We now state the three main results of our study in increasing generality that relate the convergence of Markov chain transition probabilities to the four sets of assumptions defined above.

Theorem 2.16. *A_1, B_1, C_1, $\hat{C}_1$, and P_1 are equivalent and $C_1' \Rightarrow \mathring{C}_1 \Rightarrow G_1 \Rightarrow A_1$. If $(E, \mathcal{E})$ is Borel, then all these conditions are equivalent.*

Theorem 2.17. *A_2, B_2, C_2, G_2, and P_2 are equivalent and are implied by C_2'. If $(E, \mathcal{E})$ is Borel, then each of the equivalent conditions implies C_2'.*

Theorem 2.18. *A_3, A_3', B_3, C_3, G_3, and P_3 are equivalent and are implied by C_3'. If $(E, \mathcal{E})$ is Borel, then each of the equivalent conditions implies C_3'.*

Remark 2.19. We do not know if the equivalence of all conditions with the same index holds even under our general conditions on the space $(E, \mathcal{E})$. We will comment on this in Remark 5.8.

3   First results and the proof of Theorem 2.16

Let us first state those implications in the theorems which are obvious from the definitions or are well-known.

Proposition 3.1. *We have*

a) $B_1 \Rightarrow P_1$,

b) $C_1' \Rightarrow \mathring{C}_1 \Rightarrow G_1$, $\quad P_1 \Rightarrow \hat{C}_1 \Rightarrow C_1 \Rightarrow A_1$,

c) $C_2' \Rightarrow C_2 \Rightarrow G_2 \Rightarrow A_2$, $\quad P_2 \Rightarrow A_2 \Leftrightarrow C_2$,

d) $P_3 \Rightarrow A_3' \Rightarrow A_3$, $\quad C_3' \Rightarrow C_3 \Leftrightarrow A_3$, *and* $C_3 \Rightarrow G_3 \Rightarrow A_3$.

Proof. Statement a) is a classical result and a proof can be found for example in [12, p. 328]. The remaining implications are either obvious or easy consequences of the coupling equality stated in the introduction. If, for example, A_2 holds and $x, y \in E$, then there exists some $n \in \mathbb{N}$ such that $P_n(x, \cdot) \not\perp P_n(y, \cdot)$. Hence $d(P_n(x, \cdot), P_n(y, \cdot)) < 1$ and C_2 follows from the coupling equality. $\qquad\square$

We continue by providing a slightly generalized version of the *Recurrence Lemma* from [10, Lemma 2] that will turn out to be useful later.

Lemma 3.2 (Recurrence Lemma)**.** *Assume that P satisfies Assumption A_3 or P is μ-irreducible. Then for any $B \in \mathcal{E}$ with $\mu(B) > 0$, for μ-almost every $x \in E$*

$$Q(x, B) = 1. \tag{3.1}$$

If, moreover, P satisfies Assumption A_2, then

$$Q(x, B) > 0$$

holds for every $x \in E$.

If, moreover, P satisfies Assumption A_1, then (3.1) *holds for* every $x \in E$.

Proof. For $B \in \mathcal{E}$ with $\mu(B) > 0$ define $\psi(x) := Q(x, B) = \mathbb{P}_x(X_k \in B \text{ infinitely often})$, $x \in E$. Starting X_0 with law μ, we see that $\psi(X_n)$, $n \in \mathbb{N}_0$ is a stationary process and, for $n \in \mathbb{N}$, $\psi(X_n) = \mathbb{P}(X_k \in B \text{ i.o.}|\mathcal{F}_n)$ almost surely, by the Markov property, where $\mathcal{F}_n = \sigma(X_0, ..., X_n)$ and hence $\psi(X_n) \to \mathbb{1}_{\{X_k \in B \text{ i.o.}\}}$, almost surely, by Lévy's zero-one law. Therefore $\psi(x) \in \{0, 1\}$ for μ-almost all $x \in E$. Let $\Psi_i = \{x : \psi(x) = i\}$, $i \in \{0, 1\}$. Then, by the martingale property, $P_n(x, \Psi_i) = 1$ for all $n \in \mathbb{N}_0$ and for μ-almost all $x \in \Psi_i$, $i \in \{0, 1\}$. If A_3 holds, or P is μ- irreducible, then (at least) one of the sets Ψ_0, Ψ_1 has μ-measure zero. Since $\mu(B) > 0$, Birkhoff's ergodic theorem implies $\mu(\Psi_1) > 0$, so $\mu(\Psi_0) = 0$ and $\mu(\Psi_1) = 1$, finishing the proof of the first statement.

Let Assumption A_2 hold and fix $x \in E$. Since $P_n(y, \Psi_1) = 1$ for μ-almost all y and all $n \in \mathbb{N}_0$, there exists some $y_0 \in E$ such that $P_n(y_0, \Psi_1) = 1$ for all $n \in \mathbb{N}_0$. Now A_2 applied to x and y_0 shows that there exists some $n \in \mathbb{N}$ such that $P_n(x, \Psi_1) > 0$, finishing the proof of the second claim.

Let Assumption A_1 hold and fix $x \in E$. As above, there exists some $y_0 \in E$ such that $P_n(y_0, \Psi_1) = 1$ for all $n \in \mathbb{N}_0$. Now A_1 applied to x and y_0 shows that $\lim_{n\to\infty} P_n(x, \Psi_1) = 1$, so $x \in \Psi_1$ and therefore (3.1) holds. $\qquad\square$

Proposition 3.3. *A Markov kernel P which satisfies Assumption A_3 is aperiodic.*

Proof. Suppose P has period $d \geq 2$, and let $E_1, E_2, ..., E_d$ be as in Definition 2.8. Then $\mu(E_i) > 0$ for $i = 1, 2, ..., d$. Choose $x \in E_1$, $y \in E_2$, and $n \in \mathbb{N}$ arbitrarily. Then $P_n(x, E_{n+1 (\mathrm{mod}\, d)}) = 1$ and $P_n(y, E_{n+2 (\mathrm{mod}\, d)}) = 1$ and therefore $P_n(x, \cdot) \perp P_n(y, \cdot)$. This contradicts Assumption A_3 since $(\mu \otimes \mu)(E_1 \times E_2) > 0$. $\qquad\square$

Corollary 3.4. $A_1 \Rightarrow B_1$, $A_2 \Rightarrow B_2$, *and* $A_3' \Rightarrow B_3$.

Proof. Lemma 3.2, Proposition 3.3 and Remark 2.10 immediately imply the first two implications (with $\phi := \mu$) but not the last one since the conclusion of the Recurrence Lemma under the assumption A_3 (or the stronger assumption A_3') is weaker than weak irreducibility (the exceptional sets of μ-measure 0 may depend on the set B and there may be uncountably many such sets). Therefore, we argue as follows: for $x \in E$, let $R_x := \{y \in E : x, y$ are asymptotically equivalent$\}$. Assumption A_3' and Lemma A.7 imply that $R_x \in \mathcal{E}$ and $\mu(R_x) = 1$ for μ-almost all $x \in E$. Fix $x \in E$ such that $\mu(R_x) = 1$. Since asymptotic equivalence is an equivalence relation by Remark 2.5, it follows that property A_1 holds on R_x. Using Lemma A.4, we see that B_1 holds on R_x and hence B_3 holds on E. $\qquad\square$

Before we step into the proofs of Theorems 2.16, 2.17, and 2.18, we sketch how one can see that A_i implies C_i' for $i \in \{1, 2, 3\}$. The proofs are largely identical to those in [10] where the implications $\tilde{A}_1 \Rightarrow P_1$, $A_2 \Rightarrow \tilde{P}_2$, and $A_3 \Rightarrow P_3$ were shown (with $\tilde{A}_1$ slightly stronger than A_1 and $\tilde{P}_2$ slightly weaker than P_2 and without the assumptions that the state space is Borel). We will need the Borel property only at the end of the proof when we apply the gluing lemma (Lemma A.3).

Proposition 3.5. *We have*
$$A_3 \Rightarrow P_3.$$

Further, if $(E, \mathcal{E})$ *is Borel, then*

$$A_1 \Rightarrow C_1', A_2 \Rightarrow C_2', \text{ and } A_3 \Rightarrow C_3'.$$

Idea of the proof. Under A_3, we define for $N \in \mathbb{N}$ and $p \in (0, 1)$

$$C_{N,p} := \big\{(x,y) \in E \times E : d\big(P_N(x, .), P_N(y, .)\big) \leq 1 - p\big\}.$$

$C_{N,p} \in \mathcal{E} \otimes \mathcal{E}$ by Proposition A.6 and Assumption A_3 implies $\mu \otimes \mu(C_{N,p}) > 0$ for some N and p. Fix N and p and write $C := C_{N,p}$. Let us first assume that $N = 1$ (this is without loss of generality for proving $A_3 \Rightarrow P_3$ but not without loss of generality for proving $A_3 \Rightarrow C_3'$). In [10], the authors proceed by constructing a Markov chain (Z_n), $n \in \mathbb{N}_0$ on the product space $E \times E$, which is a coupling of two chains with Markov kernel P with transition kernel S defined as

$$S\big((x,y), .\big) := \left\{ \begin{array}{ll} Q\big((x,y), .\big) & \text{if } (x,y) \in C \\ R\big((x,y), .\big) & \text{otherwise.} \end{array} \right.$$

Here, $R\big((x,y), .\big)$ is the product of $P(x, .)$ and $P(y, .)$ and the kernel Q satisfies $Q\big((x,y), \Delta\big) = 1 - d\big(P(x, .), P(y, .)\big)$ and $Q\big((x,y), .\big)$ restricted to $(E \times E) \backslash \Delta$ is absolutely continuous with repect to the product of $P(x, .)$ and $P(y, .)$ (the fact that such a kernel Q exists is stated in [10, Lemma 1]). The idea behind the definition of the kernel S is the following: whenever the chain on $E \times E$ is in a state $(x, y) \in C$, then we try to couple the two coordinates in the next step by applying Q which maximizes the coupling probability. Otherwise, we let the two coordinates move independently until the pair hits the set C. As soon as the chain (Z_n) hits the diagonal Δ it remains in that set forever. It remains to ensure that the set C is hit infinitely many times and therefore the process (Z_n) will

almost surely eventually hit Δ. The fact that (Z_n) will hit the set C almost surely in finite time can be seen as follows: consider an *independent* coupling (W_n) of two copies of the chain. Since $\mu \otimes \mu(C) > 0$, the Recurrence Lemma shows that (W_n) will hit the set C almost surely in finite time for almost all initial conditions and even for all initial conditions if we assume A_1. Since, up to the first hitting time of the set C, the processes (W_n) and (Z_n) have the same law, (Z_n) will also hit the set C almost surely in finite time. If the coupling attempt at that time is unsuccessful, then the chain (Z_n) again performs an independent coupling up to the next hit of C, which, by the same argument (and the strong Markov property and the assumptions on the kernel Q), is an almost sure event. The constructed coupling therefore shows that C_1' holds under A_1 and both C_3' and P_3 hold under A_3. Further, under A_2, for any pair $x, y \in E$ the probability that the constructed coupling is successful, is strictly positive by the second part of the Recurrence Lemma, so C_2' holds. This proves the claims in case N in the definition of the set $C_{N,p}$ can be chosen to be 1.

Finally, we assume that $N \geq 2$. The first claim follows from the case $N = 1$ since $n \mapsto d(P_n(x, .), \mu)$ is non-increasing. To see the remaining claims, we apply the previous consideration to the skeleton chain evaluated at integer multiples of N and obtain corresponding couplings $Z_{nN} = (X_{nN}, Y_{nN})$, $n \in \mathbb{N}_0$ for the skeleton chains as above. We have to make sure that these can be appropriately interpolated between successive multiples of N. This follows from an application of the gluing lemma in the appendix (which requires the state space to be Borel) to each gap between successive multiples of N (with conditionally independent interpolations), see [14, p.43] for a similar construction (it seems that the authors forgot to mention that this construction requires the space to be Borel, see Remark 5.8). $\qquad\square$

We are now ready to prove Theorem 2.16.

Proof of Theorem 2.16. Observing Proposition 3.1, Corollary 3.4 and Proposition 3.5 the claim follows once we prove that $G_1 \Rightarrow A_1$.

$G_1 \Rightarrow A_1$: Fix a pair $(x, y) \in E \times E$. We show that x and y are asymptotically equivalent. Fix $\varepsilon > 0$. By assumption there exists some $\xi \in \check{C}(\mathbb{P}_x, \mathbb{P}_y)$ such that $\lim_{k \to \infty} \xi_k(\Delta) = 1$. Since ξ^2 and $\mathbb{P}_y$ are equivalent, we can find some $\delta > 0$ such that for every $\Gamma \in \mathcal{E}^{\otimes \mathbb{N}_0}$ satisfying $\xi^2(\Gamma) < \delta$, we have $\mathbb{P}_y(\Gamma) < \varepsilon$. Let $n_0 \in \mathbb{N}_0$ be such that $\xi_k(\Delta) > 1 - \delta$ for every $k \geq n_0$. Then, for $B \in \mathcal{E}$ and $n \geq n_0$,

$$P_n(x, B) = 0 \;\Rightarrow\; \xi_n^1(B) = 0 \;\Rightarrow\; \xi_n^2(B) < \delta \;\Rightarrow\; P_n(y, B) < \varepsilon,$$

where we used absolute continuity of ξ_n^1 with respect to $P_n(x, .)$ in the first step. Reversing the roles of x and y we get $P_n(y, B) = 0 \Rightarrow P_n(x, B) < \varepsilon$ for all $n \geq n_1$. Fix $n \geq n_0 \vee n_1$ and let $B_0 \in \mathcal{E}$ be a set which maximizes $P_n(y, B)$ among all sets $B \in \mathcal{E}$ which satisfy $P_n(x, B) = 0$ and let $C_0 \in \mathcal{E}$ be a set which maximizes $P_n(x, C)$ among all sets $C \in \mathcal{E}$ which satisfy $P_n(y, C) = 0$. Note that such sets exist: pick an increasing sequence of sets $B_m \in \mathcal{E}$ such that $P_n(x, B_m) = 0$ and $P_n(y, B_m) \geq \sup\{P_n(y, B) : P_n(x, B) = 0\} - \frac{1}{m}$. Then $B_0 := \cup_m B_m$ does the job (and similarly for C_0). Define $A := E \setminus (B_0 \cup C_0)$. Then $P_n(x, A) \geq 1 - \varepsilon$, $P_n(y, A) \geq 1 - \varepsilon$ and the restrictions of $P_n(x, .)$ and $P_n(x, .)$ to A are equivalent. The claim follows since $\varepsilon > 0$ was arbitrary. $\qquad\square$

4 Proofs of Theorems 2.17 and 2.18

Building on the results from the previous section we can now prove Theorem 2.17 and 2.18.

Proof of Theorem 2.17. Thanks to Proposition 3.1, Corollary 3.4 and Proposition 3.5, the theorem is proved once we establish $B_2 \Rightarrow P_2$. Rather than adapting the proof of

$B_1 \Rightarrow P_1$ we prefer to argue along the following lines: if B_2 holds, then we show that there exists an invariant set $E_0 \subset E$ (i.e. $E_0 \in \mathcal{E}$ and $P(x, E_0) = 1$ for all $x \in E_0$) of full μ-measure on which B_1 holds and hence, by Theorem 2.16, P_1 holds. Then we show that P_2 holds on the full space E.

$B_2 \Rightarrow P_2$: We are not aware of a simple direct proof that there exists a subset of full μ-measure on which B_1 holds. Even though, thanks to Proposition A.1, irreducibility implies μ-irreducibility which, by the Recurrence Lemma, implies that $Q(x, B) = 1$ for every $B \in \mathcal{E}$ for which $\mu(B) > 0$ and μ-almost every $x \in E$, the exceptional sets may depend on B and there are (typically) uncountably many such sets B.

Since P is μ-irreducible, Proposition A.2 shows that there exists a small set $C \in \mathcal{E}$ (with ν and m as stated there). We can and will assume that $\nu(E \backslash C) = 0$. Define $G := \{x \in E : Q(x, C) = 1\}$. Then $G \in \mathcal{E}$, G is invariant, and, by the Recurrence Lemma, $\mu(G) = 1$. We claim that property B_1 holds on G. All we have to show is that $Q(x, B) = 1$ for all $x \in G$ and all $B \in \mathcal{E}$ such that $\mu(B) > 0$. Fix such a set B and let $H_{\varepsilon,l} := \{x \in G \cap C : \mathbb{P}_x\big(\bigcup_{i=0}^{l}\{X_i \in B\}\big) \geq \varepsilon\}$ for $l \in \mathbb{N}_0$ and $\varepsilon > 0$. Since $L(x, B) > 0$ for all x, there exist $l \in \mathbb{N}_0$ and $\varepsilon > 0$ such that for $H := H_{\varepsilon,l}$ we have $\nu(H) > 0$. Then $\mathbb{P}_m(y, H) \geq \nu(H) > 0$ for every $y \in C$.

This means that whenever the chain is in the set C then the probability of hitting the set B within the next $m + l$ steps is bounded away from 0. Therefore, by the strong Markov property, the chain starting in $y \in G$ will almost surely hit B infinitely often.

Using Lemma A.4, G equipped with the trace σ-field satisfies our assumption on the state space and we see that property B_1 holds on G.

Theorem 2.16 shows that property P_1 holds on G. Then, clearly, property P_3 holds on E. Since P is irreducible, we have $L(x, G) > 0$ and hence $\lim_{n \to \infty} d(P_n(x, .), \mu) < 1$ for every $x \in E$ and therefore P_2 holds on E. $\qquad \square$

Proof of Theorem 2.18. By Proposition 3.1, Corollary 3.4 and Proposition 3.5 it suffices to show that $B_3 \Rightarrow P_3$.

$B_3 \Rightarrow P_3$: We can argue like in the proof of $B_2 \Rightarrow P_2$ (the present argument is even easier). Using the very definition of weak irreducibility, we find an invariant set E_0 of full μ-measure on which B_2 and hence, using Theorem 2.17, P_2 hold. Therefore, P_3 holds on E. $\qquad \square$

5 Complements, examples, and open problems

Remark 5.1. We show that Property P_2 implies uniqueness of μ (as claimed in Remark 2.3): assume that μ and $\tilde{\mu}$ are different ipm's and let $\hat{\mu} := \frac{1}{2}\big(\mu + \tilde{\mu}\big)$. Since $P_2 \Leftrightarrow A_2$ and property A_2 is independent of the chosen ipm, we see that P_2 holds with respect to both μ and $\hat{\mu}$, so $P_n(x, .)$ converges to μ for μ-almost all x and to $\hat{\mu}$ for $\hat{\mu}$-almost all x. Since $\hat{\mu} \ll \mu$ and $\hat{\mu} \neq \mu$ this is a contradiction (this proof is adapted from [10, Proof of Corollary 1]).

Example 5.2. Let $E := \{0, 1\}$ and $P(0, \{1\}) = P(1, \{0\}) = 1$. Then the unique invariant probability measure μ is given by $\mu(\{0\}) = \mu(\{1\}) = 1/2$. For this example, the second part of property P_2 holds but the first part doesn't, so the first part of P_2 cannot be deleted without changing the class of chains for which P_2 holds.

Example 5.3. (cf. Remark 2.7.) Let $E := \mathbb{N}_0$ with the discrete σ-field $\mathcal{E}$. Define $P(x, \{x - 1\}) = 1$ for $x \geq 2$, $P(1, \{1\}) = 1$ and $P(0, \{x\}) = 2^{-x}$ for $x \in \mathbb{N}$. Clearly all transition probabilities converge to $\mu = \delta_1$ but $P_n(0, .)$ and $P_n(1, .)$ are non-equivalent for every $n \in \mathbb{N}$ (but the states 0 and 1 are asymptotically equivalent).

Example 5.4. (cf. [11, Example 5].) Let $E := \mathbb{N}_0$ with the discrete σ-field $\mathcal{E}$. Define $P(0, \{0\}) = 1$ and $P(x, \{x - 1\}) = 1/3$ and $P(x, \{x + 1\}) = 2/3$ for $x \in \mathbb{N}$. Clearly, $\mu = \delta_0$ is

the unique invariant probability measure and $P_n(x,.)$ does not converge to μ if $x > 0$, so P satisfies P_2 but not P_1. Note that for each $x, y \in E$ and $k \geq x \wedge y$, $\zeta := \delta_0 \otimes \delta_0$ satisfies $\zeta \in \tilde{C}(P_k(x,.), P_k(y,.))$ and $\zeta(\Delta) = 1$, showing that if "> 0" in Assumption G_2 is replaced by "$= 1$", then the condition does *not* imply G_1.

Remark 5.5. Note that Assumption G_1 is formally weaker than requiring that for each pair $(x,y) \in E \times E$ there exists some $\xi \in \tilde{C}(\mathbb{P}_x, \mathbb{P}_y)$ such that $\xi^1 \sim \mathbb{P}_x$ and $\xi^2 \sim \mathbb{P}_y$, but these two conditions are in fact equivalent: according to G_1 we find, for each pair (x,y), some $\check{\xi} \in \check{C}(\mathbb{P}_x, \mathbb{P}_y)$ such that $\lim_{k \to \infty} \check{\xi}_k(\Delta) = 1$ and some $\hat{\xi} \in \check{C}(\mathbb{P}_y, \mathbb{P}_x)$ such that $\lim_{k \to \infty} \hat{\xi}_k(\Delta) = 1$. Then $\xi := \frac{1}{2}\check{\xi} + \frac{1}{2}\hat{\xi}$ satisfies the formally stronger condition.

Remark 5.6. One may ask whether it is sufficient for P_1 to hold if for each pair $(x,y) \in E \times E$ and each $k \in \mathbb{N}_0$ there exists some probability measure ζ_k on $(E \times E, \mathcal{E} \otimes \mathcal{E})$ whose marginals are equivalent to $P_n(x,.)$ and $P_n(y,.)$ respectively, such that $\lim_{n \to \infty} \zeta_k(\Delta) = 1$. Again, Example 5.4 provides a negative answer. Consider ξ as in the previous example. Then $\lim_{k \to \infty} \xi_k(\Delta) \geq \lim_{k \to \infty} \xi_k(\{(0,0)\}) = 1$. Note that the marginals of the measures ξ_k are equivalent to $P_k(x,.)$ and $P_k(y,.)$ respectively but that ξ^1 and ξ^2 are not equivalent to $\mathbb{P}_x$ respectively $\mathbb{P}_y$.

Remark 5.7. From Theorem 2.16 we know that $C_1 \Rightarrow P_1$ holds since $C_1 \Rightarrow A_1 \Rightarrow B_1 \Rightarrow P_1$. Here we present an essentially well-known direct proof. For $x \in E$, $n \in \mathbb{N}$, and $A \in \mathcal{E}$ we have

$$|\mu(A) - P_n(x, A)| = \left| \int_E P_n(y, A) \, \mathrm{d}\mu(y) - P_n(x, A) \right| = \left| \int_E \Big(P_n(y, A) - P_n(x, A) \Big) \, \mathrm{d}\mu(y) \right|$$
$$\leq \int_E \left| P_n(y, A) - P_n(x, A) \right| \mathrm{d}\mu(y) \leq \int_E d\Big(P_n(y,.), P_n(x,.) \Big) \, \mathrm{d}\mu(y)$$

which converges to 0 by dominated convergence (note that Proposition A.6 shows that the last integrand is measurable with respect to y), so the claim follows.

In fact, a slight modification of the proof shows the result without employing Proposition A.6 (and without assuming that $\mathcal{E}$ is countably generated):

fix x and let $R_n(y, A) := \left| P_n(y, A) - P_n(x, A) \right|$, $n \in \mathbb{N}$. There exist sets $A_n \in \mathcal{E}$ such that

$$U_n := \sup_{A \in \mathcal{E}} \Big(\int_E R_n(y, A) \, \mathrm{d}\mu(y) \Big) \leq \int_E R_n(y, A_n) \, \mathrm{d}\mu(y) + 2^{-n},$$

which converges to 0 as $n \to \infty$ by dominated convergence.

Remark 5.8. It seems to be an open question whether all properties stated in Theorem 2.16 are equivalent even in the case in which $(E, \mathcal{E})$ is not Borel (and similarly for Theorems 2.17 and 2.18). The present proof which is based on the gluing lemma A.3 can not be applied in this case: [1] contains an example of a separable and metric space equipped with its Borel σ-field for which the conclusion in the gluing lemma fails.

Remark 5.9. It maybe of interest to generalize some of our results from Markov chains to stochastic recursive sequences driven by a stationary sequence. Such models have been investigated, for example, in [7] where they were applied to some economic models.

A Auxiliary results and measurability issues

A.1 μ-irreducibility and the existence of small sets

We start with a proposition which was announced in Remark 2.10 and whose proof is inspired by that of [12, Proposition 4.2.2], see also [4, Theorem 9.2.15].

Proposition A.1. *If P is ϕ-irreducible, then P is μ-irreducible.*

Proof. Let P be ϕ-irreducible. Then $\phi \ll \mu$ (see Remark 2.10) and, due to Lebesgue's theorem, there exists a set $B \in \mathcal{E}$ such that ϕ and μ restricted to B are equivalent and $\phi(B^c) = 0$. Note that $\mu(B) > 0$. If $\mu(B^c) = 0$, then $\phi \sim \mu$ and we are done, so we assume that $\mu(B^c) > 0$. We have to show that for any measurable set $C \subset B^c$ such that $\mu(C) > 0$ we have $L(x, C) > 0$ for every $x \in E$. Fix such x and C and define the measure

$$\nu(.) := \int_B \sum_{m=1}^{\infty} 2^{-m} P_m(y, .) \, \mathrm{d}\mu(y).$$

Invariance of μ implies $\nu \ll \mu$ and that the restriction of both measures to B are equivalent. Let $G \in \mathcal{E}$ be a set such that $\nu \sim \mu$ on G, $\nu(G^c) = 0$ and $B \subset G$.

First, we assume that $\mu(G^c) > 0$. Let $m_0 \in \mathbb{N}$ be such that $\int_{G^c} P_{m_0}(y, G) \, \mathrm{d}\mu(y) > 0$ (such an m_0 exists since P is ϕ-irreducible). Using invariance of μ, we obtain

$$\int_G P_{m_0}(y, G^c) \, \mathrm{d}\mu(y) = \int_{G^c} P_{m_0}(y, G) \, \mathrm{d}\mu(y) > 0.$$

Therefore, there exists some $\varepsilon_1 > 0$ such that for $D := \{y \in G : P_{m_0}(y, G^c) \geq \varepsilon_1\}$, we have $\mu(D) > 0$ and hence $\nu(D) > 0$, which means that there exists some $m_1 \in \mathbb{N}$ such that $\int_B P_{m_1}(y, D) \, \mathrm{d}\mu(y) > 0$.

Therefore,

$$\begin{aligned}
\nu(G^c) &\geq \int_B 2^{-m_0-m_1} P_{m_0+m_1}(y, G^c) \, \mathrm{d}\mu(y) \\
&\geq 2^{-m_0-m_1} \int_B \int_D P_{m_0}(z, G^c) P_{m_1}(y, \mathrm{d}z) \, \mathrm{d}\mu(y) \\
&\geq 2^{-m_0-m_1} \varepsilon_1 \int_B P_{m_1}(y, D) \, \mathrm{d}\mu(y) > 0,
\end{aligned}$$

contradicting the definition of G, so $\mu(G^c) = 0$.

In this case $\mu \sim \nu$ and so $\nu(C) > 0$ which implies that there exist some $\varepsilon_2 > 0$ and $m_2 \in \mathbb{N}$ such that $\tilde{D} := \{y \in B : P_{m_2}(y, C) \geq \varepsilon_2\}$ satisfies $\mu(\tilde{D}) > 0$. Next, ϕ-irreducibility and the definition of the set B imply $L(x, \tilde{D}) > 0$, which, together with the definition of $\tilde{D}$, implies $L(x, C) > 0$, so the proof of the proposition is complete. $\square$

The following proposition is an easy consequence of the rather deep Theorem 5.2.2 in [12] (which is a key step in the proof of $B_1 \Rightarrow P_1$ (in our notation)) and of the (not so deep) previous proposition.

Proposition A.2. *([12, Theorem 5.2.2]) Let P be irreducible. Then there exists a* small *set C, i.e. a set $C \in \mathcal{E}$ such that $\mu(C) > 0$ for which there exist a finite measure ν and some $m \in \mathbb{N}$ such that $\nu(C) > 0$ and $P_m(x, B) \geq \nu(B)$ for all $x \in C$ and $B \in \mathcal{E}$.*

Proof. Theorem 5.2.2 in [12] assumes that P is ψ-irreducible where ψ is a *maximal irreducibility measure*. By the previous proposition we can take $\psi = \mu$ and therefore the conclusions of [12, Theorem 5.2.2] and of Proposition A.2 are the same. $\square$

A.2 A gluing lemma

A proof of the following *gluing lemma* can be found in [1, Lemma 4.] or in [9, Lemma 4.3.2]. The conditions in [1, Lemma 4.] are even slightly weaker than ours.

Lemma A.3. *Let $(E_i, \mathcal{E}_i)$, $i = 1, 2, 3$ be Borel spaces and let ρ_1 and ρ_3 be probability measures on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ and $(E_2 \times E_3, \mathcal{E}_2 \otimes \mathcal{E}_3)$ respectively such that $\rho_1(E_1 \times B) = \rho_3(B \times E_3)$ for all $B \in \mathcal{E}_2$. Then there exists a probability measure μ on $(E_1 \times E_2 \times E_3, \mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3)$ such that $\mu(A \times E_3) = \rho_1(A)$ for all $A \in \mathcal{E}_1 \otimes \mathcal{E}_2$ and $\mu(E_1 \times B) = \rho_3(B)$ for all $B \in \mathcal{E}_2 \otimes \mathcal{E}_3$.*

A.3 Measurability issues

Lemma A.4. *Let $\tilde{E} \in \mathcal{E}$ satisfy $\mu(\tilde{E}) = 1$. Then there exists a set $\hat{E} \subset \tilde{E}$ in $\mathcal{E}$ such that $P(x, \hat{E}) = 1$ for all $x \in \hat{E}$ and $\mu(\hat{E}) = 1$. Further, for any $\tilde{E} \in \mathcal{E}$, $\tilde{E}$ equipped with the trace σ-field of $\mathcal{E}$ satisfies our basic assumptions (countably generated σ-field and measurable diagonal).*

Proof. The last statement is clear. To see the first, define $E_0 := \tilde{E}$ and $E_{i+1} := \{x \in E_i : P(x, E_i) = 1\}$, $i \in \mathbb{N}_0$. Then $\hat{E} := \bigcap_i E_i$ does the job. $\qquad\square$

In the following two statements we assume that $(E, \mathcal{E})$ satisfies our general assumptions spelled out in the introduction and that Q and $\tilde{Q}$ are Markov kernels on E.

Lemma A.5. *[9, p. 30f.] Let $\Lambda(x, y; \mathrm{d}z) := \frac{1}{2}\big(Q(x, \mathrm{d}z) + \tilde{Q}(y, \mathrm{d}z)\big)$. There exist measurable maps f and $\tilde{f}$ such that for each $A \in \mathcal{E}$, we have*

$$Q(x, A) = \int_A f(x, y; z)\,\Lambda(x, y, \mathrm{d}z), \quad \tilde{Q}(y, A) = \int_A \tilde{f}(x, y; z)\,\Lambda(x, y, \mathrm{d}z).$$

This lemma is used in [9] to prove a result which, in particular, implies the following proposition (which is not immediate since the supremum of an uncountable family of real-valued measurable functions need not be measurable).

Proposition A.6. *[9, Theorem 2.2.4 (i)] The function*

$$(x, y) \mapsto d\big(Q(x, .), \tilde{Q}(y, .)\big)$$

is measurable.

Lemma A.7. *The set of all $(x, y) \in E \times E$ for which x and y are asymptotically equivalent is a measurable subset of $(E \times E, \mathcal{E} \otimes \mathcal{E})$.*

Proof. Applying Lemma A.5 with $Q = \tilde{Q} = P_n$ we see that there exists a jointly measurable function f_n such that

$$P_n(x, A) = \int_A f_n(x, y; z)\Lambda_n(x, y; \mathrm{d}z), \ P_n(y, A) = \int_A f_n(y, x; z)\Lambda_n(x, y; \mathrm{d}z),$$

for all $x, y \in E$ (with Λ_n defined as in Lemma A.5). Defining $A_n(x, y) := \{z \in E : f_n(x, y; z)f_n(y, x; z) > 0\}$, we see that $A_n(x, y) \in \mathcal{E}$ and that $P_n(x, .)$ and $P_n(y, .)$ restricted to $A_n(x, y)$ are equivalent. Further, $A_n(x, y)$ is the largest set (up to sets of measure 0 with respect to $\Lambda_n(x, y; .)$) with this property. Observe that the map $(x, y) \mapsto P_n(x, A_n(x, y)) = \int 1_{A_n(x,y)}(z)\, P_n(x, \mathrm{d}z)$ is measurable (by a well-known application of the monotone class theorem) since the integrand is jointly measurable. The claim follows since x and y are asymptotically equivalent iff $\lim_{n\to\infty} P_n(x, A_n(x, y)) = \lim_{n\to\infty} P_n(y, A_n(x, y)) = 1$. $\qquad\square$

References

[1] P. Berti, L. Pratelli, and P. Rigo, Gluing lemmas and Skorohod representations, *Electronic Comm. Probab.* 20 (2015) 1-11. MR-3374303

[2] O. Butkovsky, A. Kulik, and M. Scheutzow, Generalized couplings and ergodic rates for SPDEs and other Markov models, *Ann. Appl. Probab.* 30 (2020) 1-39. MR-4068305

[3] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge Univ. Press, Cambridge, 1996. MR-1417491

[4] R. Douc, E. Moulines, P. Priouret, and P. Soulier, *Markov Chains*, Springer, Cham, 2018. MR-3889011

[5] J. Elstrodt, *Maß-und Integrationstheorie*, 7th edition, Springer, Berlin, 2011. MR-2257838

[6] A. Es-Sarhir, M. v. Renesse, and M. Scheutzow, Harnack inequality for functional SDEs with bounded memory, *Electronic Comm. Probab.* 14 (2009) 560-565. MR-2570679

[7] S. Foss, V. Shneer, J.P. Thomas, and T. Worrall, Stochastic stability of monotone economies in regenerative environments, *Journal of Economic Theory* 173 (2018) 334-360. MR-3737263

[8] M. Hairer and J. Mattingly, Yet another look at Harris' ergodic theorem for Markov chains, *Seminar on Stochastic Analysis, Random Fields and Applications VI*, Progr. Probab. 63, Birkhäuser, Basel, (2011) 109-117. MR-2857021

[9] A. Kulik, *Ergodic Behavior of Markov Processes*, de Gruyter, Berlin, 2018. MR-3791835

[10] A. Kulik and M. Scheutzow, A coupling approach to Doob's theorem, *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 26 (2015) 83-92. MR-3345324

[11] A. Kulik and M. Scheutzow, Generalized couplings and convergence of transition probabilities, *Probab. Theory Related Fields* 171 (2018) 333-376. MR-3800835

[12] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Second edition, Cambridge Univ. Press, Cambridge, 2009. MR-2509253

[13] S. Orey, *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand Reinhold, London, 1971. MR-0324774

[14] G. O. Roberts and J. S. Rosenthal, General state space Markov chains and MCMC algorithms, *Probability Surveys* 1 (2004) 20-71. MR-2095565

[15] A. Veretennikov, Coupling methods for Markov chains under integral Doeblin type conditions, *Theory Stoch. Processes* 8 (2002) 383-391. MR-2027410

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm