A review of asymptotic convergence for general state space Markov chains

by

Jeffrey S. Rosenthal*

(March, 1999; slightly revised April, 2002.)

Abstract. We review notions of small sets, ϕ-irreducibility, etc., and present a simple proof of asymptotic convergence of general state space Markov chains to their stationary distributions.

1. Introduction.

It is a standard fact (see e.g. [3], Theorem 8.6) that a countable (or finite) state space Markov chain, which is both irreducible and aperiodic, and which possesses a stationary distribution, will converge to this distribution as the number of iterations goes to infinity.

For Markov chains on general (uncountable) state spaces, the situation is somewhat more complicated. However, a very similar result still holds. This result (stated as Theorem 1 below) was proved and refined over the years by Doeblin [5], Orey [14], Jain and Jameson [8], Athreya and Ney [4], Nummelin [12], and others. Discussions of this result appear in Nummelin [13], Meyn and Tweedie [10], Tierney [20], Smith and Roberts [19], Asmussen [1], Athreya et al. [2], and elsewhere. Some of these discussions were motivated partially by Markov chain Monte Carlo algorithms, which have recently received a great deal of attention (see e.g. [6], [19], [20], [7], [16]), and for which convergence issues are extremely important.

---

* Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: `jeff@math.toronto.edu`. Web: `http://markov.utstat.toronto.edu/jeff/`.
Supported in part by NSERC of Canada.

Despite the numerous references to the general state space convergence result, it is perhaps still true that there is no simple, brief proof available which explains the convergence in intuitive probabilistic terms. This note attempts to answer that need. We provide a simple and intuitive proof of the general state space result. Our proof is entirely self-contained with one exception: we quote without proof the great result of Jain and Jameson [8] about the existence of small sets. We also discuss some related issues such as quantitative convergence bounds (cf. [11], [17], [18]). This paper is purely expository; no new results are presented.

2. Definitions and theorem statement.

We review some standard Markov chain terminology; for further details see e.g. [10].

Let $\mathcal{X}$ be a general measurable set. We consider a Markov chain on $\mathcal{X}$, governed by transition probabilities $P(x, dy)$. This means that for each $x \in \mathcal{X}$ and measurable subset $A \subseteq \mathcal{X}$, the quantity $P(x, A)$ represents the probability of jumping from x to somewhere in the set A. Formally, we assume that for each fixed $x \in \mathcal{X}$, the function $P(x, \cdot)$ is a probability measure on $\mathcal{X}$; also, for each fixed measurable set $A \subseteq \mathcal{X}$, the function $P(\cdot, A)$ is a measurable function.

In terms of $P(x, dy)$, higher-order transition probabilities can be defined, by $P^1(x, dy) = P(x, dy)$, and

$$P^{n+1}(x, A) \; = \; \int_{\mathcal{X}} P^n(x, dy) \, P(y, A) \qquad n = 1, 2, 3, \ldots .$$

Given transition probabilities $P(x, dy)$, there exist random variables $X_0, X_1, \ldots$ such that $\mathbf{P}(X_{n+1} \in A \,|\, X_n = x) = P(x, A)$ for $n = 0, 1, 2, \ldots$. We sometimes write $\mathbf{P}_x(\cdots)$ as short-hand for $\mathbf{P}(\cdots \,|\, X_0 = x)$. We define the random variable $\tau_A = \inf\{n \geq 1; \; X_n \in A\}$ (which could equal infinity) to be the first return time to A.

The Markov chain is *ϕ-irreducible* if there exists a non-zero measure ϕ on $\mathcal{X}$, such that for any measurable $A \subseteq \mathcal{X}$ with $\phi(A) > 0$, we have $\mathbf{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$; i.e., such that any set of positive ϕ measure has positive probability of being hit from any starting point x. (This is the general state space analog of irreducibility.)

Like for countable chains, a Markov chain is *aperiodic* if there does not exist a partition $\mathcal{X} = \mathcal{X}_1 \dot{\bigcup} \mathcal{X}_2 \dot{\bigcup} \ldots \dot{\bigcup} \mathcal{X}_d$ for some $d \geq 2$, such that $P(x, \mathcal{X}_{i+1}) = 1$ for all $x \in \mathcal{X}_i$ $(1 \leq i \leq d-1)$ and $P(x, \mathcal{X}_1) = 1$ for all $x \in \mathcal{X}_d$.

A probability distribution π on $\mathcal{X}$ is a *stationary distribution* for a Markov chain with transition probabilities $P(x, dy)$ if

$$\pi(A) \;=\; \int_{\mathcal{X}} \pi(dy)\, P(y, A)\,, \qquad \text{for all measurable } A \subseteq \mathcal{X}\,.$$

Finally, we recall that given two probability distributions μ and ν, their *total variation distance* is given by

$$\|\mu - \nu\| \;=\; \sup_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|\,,$$

where the supremum is taken over all measurable subsets A.

In terms of this standard terminology, we can state the result we wish to prove:

Theorem 1. *Let $P(x, dy)$ be the transition probabilities for a ϕ-irreducible, aperiodic Markov chain on a general state space $\mathcal{X}$, having stationary distribution π. Then for π-a.e. $x \in \mathcal{X}$, we have*

$$\lim_{n \to \infty} \|P^n(x, \cdot) - \pi(\cdot)\| \;=\; 0\,.$$

In words, from π-almost-every starting point x, the total variation distance of the Markov chain to the stationary distribution goes to 0 *as the number of steps goes to infinity.*

The remainder of this paper sets out to prove Theorem 1.

3. Small sets, coupling, and splittings.

Given a Markov chain with transition kernel $P(x, dy)$ on a state space $\mathcal{X}$, a subset $C \subseteq \mathcal{X}$ is said to be *small* if there is a probability measure $Q(\cdot)$ on $\mathcal{X}$, a positive integer k_0, and $\epsilon > 0$, such that

$$P^{k_0}(x, \cdot) \;\geq\; \epsilon\, Q(\cdot)\,, \qquad \text{for all } x \in C\,, \tag{1}$$

i.e. such that $P^{k_0}(x, A) \geq \epsilon\, Q(A)$ for all $x \in C$ and all measurable subsets $A \subseteq \mathcal{X}$.

One use of small sets is to construct *couplings* (see e.g. Pitman [15]; Lindvall [9]). Specifically, suppose we wish to jointly construct two copies $\{X_n\}_{n=0}^\infty$ and $\{Y_n\}_{n=0}^\infty$ of the Markov chain, i.e. two sequences of random variables such that for $n = 0, 1, 2, \ldots$,

$$\mathcal{L}(X_{n+1} \,|\, X_n) \;=\; P(X_n, \cdot)\,; \qquad \mathcal{L}(Y_{n+1} \,|\, Y_n) \;=\; P(Y_n, \cdot)\,. \tag{2}$$

Given a small set C as above, and starting values (X_0, Y_0), one way to proceed is the following "splitting construction" (see e.g. [4], [13], [10]), defined iteratively for $n = 0, 1, 2, \ldots$:

Given (X_n, Y_n),

1. If $(X_n, Y_n) \notin C \times C$, then we *independently* choose $X_{n+1} \sim P(X_n, \cdot)$ and $Y_{n+1} \sim P(Y_n, \cdot)$. We then replace n with $n+1$ and iterate.
2. If $(X_n, Y_n) \in C \times C$, then
   (a) with probability ϵ, we choose $X_{n+k_0} = Y_{n+k_0} \sim Q(\cdot)$;
   (b) with probability $1-\epsilon$, we independently choose $X_{n+k_0} \sim \frac{1}{1-\epsilon}(P^{k_0}(X_n, \cdot) - \epsilon\, Q(\cdot))$ and $Y_{n+k_0} \sim \frac{1}{1-\epsilon}(P^{k_0}(Y_n, \cdot) - \epsilon\, Q(\cdot))$.

   Under either 2(a) or 2(b), we then go back and "fill in" the missing values $\{(X_k, Y_k)\}_{n<k<n+k_0}$, conditionally independently. That is, we choose $\{X_k\}_{n<k<n+k_0}$, conditional on the chosen values of X_n and X_{n+k_0}, but conditionally independently of the values $\{Y_k\}$. We similarly choose $\{Y_k\}_{n<k<n+k_0}$, conditional on the chosen values of Y_n and Y_{n+k_0}, but conditionally independently of the values $\{X_k\}$. Under 2(b), we then replace n with $n + k_0$ and iterate.

We continue this iterative procedure until the first time we perform option 2(a). As soon as we perform option 2(a), we let $T_* = n + k_0$, and then complete the construction by choosing $X_{k+1} = Y_{k+1} \sim P(X_k, \cdot)$ for $k = T_*, T_* + 1, T_* + 2, \ldots$. Here T_* is the *coupling time*, and we shall make reference to it later in the paper.

This completes the description of the splitting construction. We note that, by (1), the quantity $\frac{1}{1-\epsilon}(P^{k_0}(X_{\ell k_0}, \cdot) - \epsilon\, Q(\cdot))$ of option 2(b) above is indeed a probability measure. Also, rules 2(a) and 2(b) have been chosen so that $\mathcal{L}(X_{n+k_0} \,|\, X_n) \;=\; P^{k_0}(X_n, \cdot)$ (and similarly for Y_{n+k_0}). [Indeed, this follows since

$$\epsilon\, Q(\cdot) + (1-\epsilon)\left[\frac{1}{1-\epsilon}(P^{k_0}(X_n,\cdot) - \epsilon\, Q(\cdot))\right] = P^{k_0}(X_n,\cdot)\,,$$

and similarly for Y_n.] It then follows that (2) always holds, i.e. that the two chains each marginally follow the original transition probabilities.

We note further that, by construction, we have $X_n = Y_n$ whenever $n \geq T_*$. The reason this is helpful is the *coupling inequality* (see e.g. Pitman [15]; Lindvall [9]), which states that

$$\|\mathcal{L}(X_n) - \mathcal{L}(Y_n)\| \;\leq\; \mathbf{P}(X_n \neq Y_n)\,.$$

Indeed, we have that

$$\begin{aligned} &|\mathbf{P}(X_n \in A) - \mathbf{P}(Y_n \in A)| \\ =\; &\big|\mathbf{P}(X_n \in A,\ X_n = Y_n) + \mathbf{P}(X_n \in A,\ X_n \neq Y_n) \\ &\quad -\mathbf{P}(Y_n \in A,\ X_n = Y_n) - \mathbf{P}(Y_n \in A,\ X_n \neq Y_n)\big| \\ =\; &|\mathbf{P}(X_n \in A,\ X_n \neq Y_n) - \mathbf{P}(Y_n \in A,\ X_n \neq Y_n)| \\ \leq\; &\mathbf{P}(X_n \neq Y_n)\,. \end{aligned}$$

[We have used that $\mathbf{P}(X_n \in A,\ X_n = Y_n) = \mathbf{P}(Y_n \in A,\ X_n = Y_n)$, and also that the difference between two non-negative quantities each $\leq \mathbf{P}(X_n \neq Y_n)$ must itself be $\leq \mathbf{P}(X_n \neq Y_n)$.]

The coupling inequality suggests that to bound convergence in total variation distance, it suffice to construct two copies of the chain which have a good chance of becoming equal. Furthermore, the above splitting construction suggests that small sets are a good way of constructing such copies. Thus, combining small sets with the coupling inequality shall be the key to proving Theorem 1.

To facilitate the proof, we require a remarkable result about small sets, due to Jain and Jameson [8] (see also Orey [14]). We shall not prove it here; for modern proofs see e.g. Nummelin [13], p. 16, or Meyn and Tweedie [10], Theorem 5.2.2. The key idea (cf. [10], Theorem 5.2.1) is to extract the part of $P^{k_0}(x,\cdot)$ which is absolutely continuous with respect to the measure ϕ, and then to find a C with $\phi(C) > 0$ such that this density part is at least $\delta > 0$ throughout C.

Theorem 2. *(Jain and Jameson [8]) Every ϕ-irreducible Markov chain contains a small set $C \subseteq \mathcal{X}$ with $\phi(C) > 0$.*

In fact more is true: every set $B \subseteq \mathcal{X}$ with $\phi(B) > 0$ in turn contains a small set $C \subseteq B$ with $\phi(C) > 0$. Furthermore, we may assume $Q(C) > 0$ (where Q is the minorising measure from the definition of small set). We do not use these facts here.

4. Return lemmas.

Before proceeding to the proof of Theorem 1, we present some lemmas related to ϕ-irreducibility. The first is not strictly necessary for our proof, but it does provide greater intuition about the nature of the irreducibility measure ϕ.

Lemma 3. *If a Markov chain is ϕ-irreducible, and has a stationary distribution π, then $\phi \ll \pi$, i.e. $\pi(A) > 0$ whenever $\phi(A) > 0$.*

Proof. Let $\phi(A) > 0$. Then $\mathbf{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$. Hence, $\mathcal{X} = \bigcup_{n,m \in \mathbf{N}} \{x \in \mathcal{X}\,;\ P^n(x, A) \geq \frac{1}{m}\}$. By countable additivity, we can find $n, m \in \mathbf{N}$ and a subset $B \subseteq \mathcal{X}$ with $\pi(B) > 0$, such that $P^n(x, A) \geq 1/m$ for all $x \in B$. But then

$$\pi(A) = \int_{\mathcal{X}} P^n(x, A)\, \pi(dx) \geq \int_B P^n(x, A)\, \pi(dx) \geq \int_B (1/m)\, \pi(dx) = \pi(B)\,/\,m > 0\,. \quad \blacksquare$$

Lemma 4. *Consider a Markov chain on a state space $\mathcal{X}$, having stationary distribution π. Suppose that for some $A \subseteq \mathcal{X}$, we have $\mathbf{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$. Then for π-almost-every $x \in \mathcal{X}$, $\mathbf{P}_x(\tau_A < \infty) = 1$.*

Proof. Suppose to the contrary that this is not true. Then by countable additivity we can find $m \in \mathbf{N}$ and a subset $B \subseteq \mathcal{X}$ with $\pi(B) > 0$, such that $\mathbf{P}_x(\tau_A < \infty) \leq 1 - \frac{1}{m}$ for all $x \in B$. On the other hand, since $\mathbf{P}_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$, we can find $n_0 \in \mathbf{N}$ and $\delta_0 > 0$ and $B' \subseteq B$ with $\pi(B') > 0$ and with $P^{n_0}(x, A) \geq \delta_0$ for all $x \in B'$.

Now, setting $\eta_{B'} = \#\{k \geq 1;\ X_{n_0 k} \in B'\}$, we have that $\mathbf{P}(\tau_A = \infty,\ \eta_{B'} = r) \leq (1 - \delta_0)^r$. In particular, $\mathbf{P}(\tau_A = \infty,\ \eta_{B'} = \infty) = 0$. Hence, for $x \in B'$, we have

$$\mathbf{P}_x(\tau_A = \infty,\ \eta_{B'} < \infty) \ = \ 1 - \mathbf{P}_x(\tau_A = \infty,\ \eta_{B'} = \infty) - \mathbf{P}_x(\tau_A < \infty)$$

$$\geq \; 1 - 0 - \left(1 - \frac{1}{m}\right) \; = \; \frac{1}{m}\,.$$

Hence, there is $\ell \in \mathbf{N}$, $\delta > 0$, and $B'' \subseteq B'$ with $\pi(B'') > 0$, such that

$$\mathbf{P}_x\left(\tau_A = \infty,\; \sup\{k \geq 1;\; X_{n_0 k} \in B'\} < \ell\right) \geq \delta\,, \qquad x \in B''\,.$$

So, letting $n = n_0 \ell$, we thus see that for any $x \in B''$, we have $\mathbf{P}_x(\tau_A = \infty,$ and $X_{kn} \notin B'$ for all $k \geq 1) \geq \delta$.

This subset B'' shall be the key to the proof. Indeed, we compute that for any $j \in \mathbf{N}$,

$$\pi(A^C) \;=\; \int_{\mathcal{X}} \pi(dy)\, P^{jn}(y, A^C)$$

$$=\; \int_{\mathcal{X}} \pi(dy)\, \mathbf{P}_y(X_{jn} \in A^C)$$

$$\geq \int_{\mathcal{X}} \pi(dy)\, \mathbf{P}_y\left(\bigcup_{i=0}^{j-1}[X_{in} \in B'',\; X_{(i+1)n} \notin B',\; X_{(i+2)n} \notin B',\; \ldots,\; X_{(j-1)n} \notin B',\; X_{jn} \in A^C]\right)$$

$$= \sum_{i=0}^{j-1} \int_{\mathcal{X}} \pi(dy) \mathbf{P}_y\left(X_{in} \in B'',\; X_{(i+1)n} \notin B',\; X_{(i+2)n} \notin B',\; \ldots,\; X_{(j-1)n} \notin B',\; X_{jn} \in A^C\right)\,.$$

$$= \sum_{i=0}^{j-1} \int_{\mathcal{X}} \pi(dy) \mathbf{P}_y\left(X_0 \in B'',\; X_n \notin B',\; X_{2n} \notin B',\; \ldots,\; X_{(j-i-1)n} \notin B',\; X_{(j-i)n} \in A^C\right)\,.$$

(We have used that the above union is disjoint, since $B'' \cup (B')^C = \emptyset$.)

But the inner integral is at least as large as the probability that, being in the stationary distribution, we begin in B'', and then have $\tau_A = \infty$ and furthermore never return to $B' \supseteq B''$ in any integer multiple of n steps. We thus have

$$\pi(A^C) \;\geq\; \sum_{i=0}^{j-1} [\pi(B'')\,\delta] \;=\; j\,\pi(B'')\,\delta\,.$$

This leads to the desired contraction: if we choose j large enough, we can make this quantity larger than 1, which is impossible.

Hence, our original assumption must be false, and we conclude that for π-almost-every $x \in \mathcal{X}$, we have $\mathbf{P}_x(\tau_A < \infty) = 1$, as claimed. ∎

Corollary 5. *Assume the hypotheses of Lemma 4. Then for π-almost-every $x \in \mathcal{X}$, we have $\mathbf{P}_x(X_n \in A \text{ i.o.}) = 1$. That is, with probability 1, the chain will hit the set A infinitely often.*

Proof. Define $r_A = \sup\{n \geq 1;\ X_n \in A\}$ (which could be infinite). Then for fixed $n \in \mathbf{N}$,

$$\pi\{x \in \mathcal{X};\ \mathbf{P}_x(r_A = n) > 0\} = \int\int \pi(dx)\, P^n(x, dy)\, \mathbf{P}_y(\tau_A = \infty)$$
$$= \int \pi(dy)\, \mathbf{P}_y(\tau_A = \infty) = 0$$

since by Lemma 4 we have $\mathbf{P}_y(\tau_A < \infty) = 1$ for π-a.e. $y \in \mathcal{X}$. But then by countable additivity, $\pi\{x \in \mathcal{X};\ \mathbf{P}_x(r_A < \infty) > 0\} = 0$. This gives the result. ∎

We also need a lemma related to aperiodicity.

Lemma 6. *Let $C \subseteq \mathcal{X}$ be a small set with $P^{k_0}(x, \cdot) \geq \epsilon\, Q(\cdot)$ for all $x \in C$. Let $S = \{n \geq 1;\ \int Q(dx)\, P^n(x, C) > 0\}$. Then there is $n_* \in \mathbf{N}$ with $S \supseteq \{n_*, n_*{+}1, n_*{+}2, \ldots\}$.*

Proof. Let $T = S + k_0 \equiv \{s + k_0;\ s \in S\}$. Then T represents the times at which it is possible for the chain, if started in initial distribution $Q(\cdot)$, to again be in the distribution $Q(\cdot)$ (having chosen option 2(a) in the splitting construction). Thus, T is *additive*, in the sense that if $n, m \in T$ then $n + m \in T$.

Furthermore, we claim that $gcd(T) = 1$. Indeed, if $gcd(T) = d > 1$, then for $1 \leq i \leq d$, let $\mathcal{X}_i = \{x \in \mathcal{X};\ \exists \ell \in \mathbf{N},\ \int Q(dx)\, P^{\ell d + i}(x, C) > 0\}$. Then $\mathcal{X} = \mathcal{X}_1 \dot{\bigcup} \ldots \dot{\bigcup} \mathcal{X}_d$ forms a partition with respect to which the Markov chain is periodic (of period d). This contradicts the assumption that the chain is aperiodic.

Hence, $gcd(T) = 1$, and T is additive. It is then a standard and easy fact (cf. [3], p. 541) that there is $n' \in \mathbf{N}$ such that $T \supseteq \{n', n'{+}1, n'{+}2, \ldots\}$. Hence, setting $n_* = n' - k_0$, we are done. ∎

5. Proof of Theorem 1.

We now proceed to the proof of Theorem 1. We suppose that $P(x, dy)$ are the transition probabilities for a ϕ-irreducible, aperiodic Markov chain on a general state space $\mathcal{X}$, having stationary distribution π.

From Theorem 2, there is a small set $C \subseteq \mathcal{X}$ with $\phi(C) > 0$, such that $P^{k_0}(x, \cdot) \geq \epsilon\, Q(\cdot)$ for all $x \in C$. (By Lemma 3, we also have $\pi(C) > 0$.)

We now choose $(X_0, Y_0) = (x, y)$, for any fixed $x, y \in \mathcal{X}$. We then jointly define $\{X_n\}_{n=0}^{\infty}$ and $\{Y_n\}_{n=0}^{\infty}$ according to the "splitting construction" of Section 3 above.

Since $\phi(C) > 0$, we have $\mathbf{P}_x(\tau_C < \infty) > 0$ and $\mathbf{P}_y(\tau_C < \infty) > 0$. Hence, there are $n_x, n_y \in \mathbf{N}$ such that $P^{n_x}(x, C) > 0$ and $P^{n_y}(y, C) > 0$. Let n_* be as in Lemma 6. Then, recalling that $P^{k_0}(x, \cdot) \geq \epsilon\, Q(\cdot)$ for all $x \in C$, we see that $P^{n_x + k_0 + n}(x, C) > 0$ whenever $n \geq n_*$, and similarly $P^{n_y + k_0 + n}(y, C) > 0$ whenever $n \geq n_*$.

Now, let $\ell \geq \max(n_x, n_y) + k_0 + n_*$. For the joint chain $\{(X_n, Y_n)\}$, we claim that $\mathbf{P}_{(x,y)}(\tau_{C \times C} \leq \ell) > 0$. Indeed, until such time as the joint chain enters $C \times C$, the joint chain behaves as two independent chains. Thus, $\mathbf{P}_{(x,y)}(\tau_{C \times C} \leq \ell) \geq P^{\ell}(x, C)\, P^{\ell}(y, C) > 0$. Hence, in particular, $\mathbf{P}_{(x,y)}(\tau_{C \times C} < \infty) > 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{X}$.

Now, let $G \subseteq \mathcal{X} \times \mathcal{X}$ be the set of (x, y) for which $\mathbf{P}_{(x,y)}\big((X_n, Y_n) \in C \times C \ i.o.\big) = 1$. We claim that

$$(\pi \times \pi)(G) = 1\,. \tag{3}$$

Indeed, by applying Lemma 4 to the joint chain, we see that the joint chain will return to $C \times C$ with probability 1 from $(\pi \times \pi)$-a.e. $(x, y) \notin C \times C$. Once the chain reaches $C \times C$, then conditional on not coupling, the chain's update distribution is absolutely continuous with respect to $P \times P$ and hence (again by Lemma 4) will return again to $C \times C$ with probability 1. Conditional on coupling, the two chains will then proceed as one (starting from the distribution $Q(\cdot)$ which must be absolutely continuous with respect to $\pi(\cdot)$), and hence will return to $C \times C$ by Lemma 4 applied to a single chain. Hence, in any case, the chain will repeatedly return to $C \times C$ with probability 1. This proves (3).

Now, each time that $(X_n, Y_n) \in C \times C$ (excluding those times that are within k_0 of a previous visit to $C \times C$), by part 2(a) of the splitting construction, we see that there is independent probability $\epsilon > 0$ that the two chains will couple k_0 iterations later (i.e., that we will then have $T_* = n + k_0$, where T_* is the coupling time from the splitting construction of Section 3). But if $(x, y) \in G$ then with probability 1 there will be infinitely many such opportunities. We conclude that

$$\lim_{n\to\infty} \mathbf{P}_{(x,y)}(T_* > n) \;=\; 0\,, \qquad (x, y) \in G\,. \tag{4}$$

Furthermore, the splitting construction ensures that $X_n = Y_n$ for all $n \geq T_*$. Hence, the coupling inequality tells us that

$$\|P^n(x,\cdot) - P^n(y,\cdot)\| \;\leq\; \mathbf{P}_{(x,y)}(X_n \neq Y_n) \;\leq\; \mathbf{P}_{(x,y)}(T_* > n)\,.$$

Now, for $x \in \mathcal{X}$ let $G_x = \{y \in \mathcal{X};\ (x, y) \in G\}$, and let $\overline{G} = \{x \in \mathcal{X};\ \pi(G_x) = 1\}$. Then for $x \in \overline{G}$, we have by (4) and the generalised triangle inequality that

$$\begin{aligned}
\lim_{n\to\infty} \|P^n(x,\cdot) - \pi\| \;&=\; \lim_{n\to\infty} \left\| \int \pi(dy) P^n(x,\cdot) - \int \pi(dy) P^n(y,\cdot) \right\| \\
&\leq\; \lim_{n\to\infty} \int \pi(dy) \|P^n(x,\cdot) - P^n(y,\cdot)\| \\
&\leq\; \lim_{n\to\infty} \int \pi(dy) \mathbf{P}_{(x,y)}(T_* > n) \;=\; 0\,.
\end{aligned} \tag{5}$$

(Here we have used the bounded convergence theorem plus the fact that, for $x \in \overline{G}$, we have $\lim_{n\to\infty} \mathbf{P}_{(x,y)}(T_* > n) = 0$ for π-a.e. $y \in \mathcal{X}$.)

Finally, we claim that $\pi(\overline{G}) = 1$. Indeed, if we had $\pi(\overline{G}) < 1$, then we would have

$$(\pi \times \pi)(G^C) = \int_{\mathcal{X}} \pi(dx) \pi(G_x^C) = \int_{\overline{G}^C} \pi(dx)[1 - \pi(G_x)] > 0\,,$$

contradicting (3). Hence, $\pi(\overline{G}) = 1$. The theorem now follows from (5). ∎

6. Extensions and remarks.

It is possible to extend this result in various ways for particular cases, by being more careful about the return times to $C \times C$ (i.e., the law of $\tau_{C \times C}$) and/or about the constant ϵ from the small set definition (1).

For example, if one proves that $\tau_{C \times C}$ has finite exponential moments, then one can prove that the convergence of Theorem 1 is exponentially quick (cf. [10], Theorem 15.0.1).

Similarly, if one provides quantitative bounds on the tail of $\tau_{C \times C}$, and on the ϵ of the small set, then one can obtain quantitative bounds on the distance to stationarity after a finite number n of iterations (cf. [11], [17], [18]).

Furthermore, note that the qualifier "for π-a.e. $x \in \mathcal{X}$" in the statement of Theorem 1 is indeed necessary and cannot be replaced by "for all $x \in \mathcal{X}$". For instance, here is a counter-example due to C. Geyer. Let $\mathcal{X} = \{0, 1, 2, \ldots\}$, with $P(0,0) = 1$, and for $n \geq 1$, $P(n, n+1) = 1 - P(n, 0) = 1 - \frac{1}{n^2}$. Then the stationary distribution π is a point-mass at 0, and the chain is indeed π-irreducible and aperiodic. However, if we begin at any point in the set $\{2, 3, 4, \ldots\}$ (which has π-measure 0), then there is positive probability that we will continue moving from n to $n+1$, and never settle down at the point 0.

Of course, if $P(x, \cdot) \ll \pi(\cdot)$ for all $x \in \mathcal{X}$, then we can indeed replace "for π-a.e. $x \in \mathcal{X}$" by "for all $x \in \mathcal{X}$". More generally, we can make this replacement if the chain is *Harris recurrent*; see e.g. [20] for discussion of this.

Finally, we note that small sets may also be thought of as providing "regeneration events", whereby with probability ϵ a chain "regenerates" by jumping into the fixed distribution $Q(\cdot)$. This reduces convergence of the chain to convergence of the "time since last regeneration" process. This approach eliminates the need to consider a second chain $\{Y_n\}$, thus simplifying some arguments. However, it does require some renewal theory to study the time since the last regeneration. See e.g. Asmussen [1] for a good discussion of this approach.

REFERENCES

[1] S. Asmussen (1987), Applied Probability and Queues. John Wiley & Sons, New York.

[2] K.B. Athreya, H. Doss, and J. Sethuraman (1996), On the Convergence of the Markov Chain Simulation Method. Ann. Stat. **24**, 69–100.

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

[4] K.B. Athreya and P. Ney (1978), A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. **245**, 493-501.

[5] W. Doeblin (1940), Eléments d'une théorie générale des chaînes simples constantes de Markov. Annales Scientifiques de l'Ecolre Normale Supérieure **57 (III)**, 61–111.

[6] A.E. Gelfand and A.F.M. Smith (1990), Sampling based approaches to calculating marginal densities. J. Amer. Stat. Assoc. **85**, 398-409.

[7] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, ed. (1996), Markov chain Monte Carlo in practice. Chapman and Hall, London.

[8] N. Jain and B. Jamison (1967), Contributions to Doeblin's theory of Markov processes. Z. Wahrsch. Verw. Geb. **8**, 19–40.

[9] T. Lindvall (1992), Lectures on the Coupling Method. Wiley & Sons, New York.

[10] S.P. Meyn and R.L. Tweedie (1993), Markov chains and stochastic stability. Springer-Verlag, London.

[11] S.P. Meyn and R.L. Tweedie (1994), Computable bounds for convergence rates of Markov chains. Ann. Appl. Prob. **4**, 981-1011.

[12] E. Nummelin (1978), Uniform and ratio limit theorems for Markov renewal and semi-regenerative processes on a general state space. Ann. Inst. Henri Poincaré Series B **14**, 119–143.

[13] E. Nummelin (1984), General irreducible Markov chains and non-negative operators. Cambridge University Press.

[14] S. Orey (1971), Lecture notes on limit theorems for Markov chain transition probabilities. Van Nostrand Reinhold, London.

[15] J.W. Pitman (1976), On coupling of Markov chains. Z. Wahrsch. verw. Gebiete **35**, 315-322.

[16] G.O. Roberts and J.S. Rosenthal (1998), Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). Canadian J. Stat. **26**, 5–31.

[17] J.S. Rosenthal (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558-566.

[18] J.S. Rosenthal (1996), Convergence of Gibbs sampler for a model related to James-Stein estimators. Stat. and Comput. **6**, 269-275.

[19] A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3-24.

[20] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701-1762.