Z. Wahrscheinlichkeitstheorie verw. Gebiete 31, 95–106 (1975)

A Maximal Coupling for Markov Chains

David Griffeath

0. Introduction

A homogeneous Markov chain X_n with denumerable state space S and n-step transition matrix $p^n=(p_{ik}^{(n)})_{i,k\in S}$ is called *weakly ergodic* if

$$\lim_{n\to\infty}\sum_k |p_{ik}^{(n)}-p_{jk}^{(n)}|=0 \qquad \text{for all } i,j\in S.$$

This property expresses asymptotic "loss of memory", a notion with various equivalent formulations: trivial tail field, absence of non-constant space–time harmonic functions, and mixing, in particular. Over recent years, a "coupling method" for proving ergodicity results has been developed by Vasershtein [12], and many others. The technique involves constructing two copies of X_n which start from different states and evolve simultaneously in such a way that if they ever reach the same state, then they are "pasted together" from that time on. More precisely, a coupling is a bivariate process $\tilde{X}_n=(X_n^1, X_n^2)$ on the state space $\tilde{S}=S\times S$ such that if $\tilde{X}_n=(k,k)$ for some N, then $\tilde{X}_n$ remains on the diagonal $D=\{(s,s); s\in S\}$ of $\tilde{S}$ for all $n>N$. The basic coupling result states that if $\tilde{X}_n$ reaches the diagonal, starting from any $(i,j)\in\tilde{S}$, in a finite time with probability one (in this case we call $\tilde{X}_n$ *successful*), then the original chain X_n is weakly ergodic. Coupling has been used most extensively in the study of Markovian lattice interactions [3, 7, 12], but may also be used to derive ergodicity criteria for homogeneous and nonhomogeneous Markov chains [6, 11]. In this paper we will briefly survey some of the known applications of couplings to Markov chains, and then prove a new theorem which states that any weakly ergodic chain has a successful "maximal" coupling $\tilde{X}_n$. Using this maximal coupling, it is possible to obtain a necessary and sufficient condition for weak ergodicity of an arbitrary Markov chain. Our construction also gives additional insight into the original proof of the weak ergodic theorem for irreducible aperiodic recurrent chains, due to Orey [8], as well as the corresponding theorem for aperiodic random walk.

1. Preliminaries

Let X_n be a (homogeneous) Markov chain with time parameter set $\mathbb{N}=\{0,1,\ldots\}$ and denumerable state space S. Let $\Omega=S^{\mathbb{N}}$, and let $\omega=(\omega_0,\omega_1,\ldots)$ be the coordinate process on Ω for X_n. Denote by $p=(p_{ik})_{i,k\in S}$ the 1-step transition matrix for X_n, and let P_i be the measure induced on $(\Omega,\mathscr{B})$ when X_n starts in state i. Here $\mathscr{B}=\sigma[(\omega_n)_{n\in\mathbb{N}}]$ is the σ-algebra generated by the coordinate functions. Also, write $p_{ik}^{(n)}=P_i(\omega_n=k)$. Similarly, let P_ν be the measure on $(\Omega,\mathscr{B})$ when X_n has starting probability measure ν.

Denote $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, $a^+ = a \vee 0$; $a, b \in \mathbb{R}$. Now define

$$\delta_{ij}^{(n)} = \tfrac{1}{2} \sum_{k \in S} |p_{ik}^{(n)} - p_{jk}^{(n)}|, \quad \alpha_{ik}^{(n)} = \sum_{k \in S} (p_{ik}^{(n)} \wedge p_{jk}^{(n)}); \quad i, j \in S, \ n \in \mathbb{N}.$$

Recalling that for any reals a, b we have $a \wedge b = \frac{1}{2}(a + b - |a - b|)$, it is easy to check that $\alpha_{ij}^{(n)} = 1 - \delta_{ij}^{(n)}$. The Markov chain X_n is *weakly ergodic* iff

$$\lim_{n \to \infty} \delta_{ij}^{(n)} = 0 \qquad \text{for all } i, j \in S. \tag{1}$$

One can verify that $\delta_{ij}^{(n)} = \sup_{E \subset S} |P_i(\omega_n \in E) - P_j(\omega_n \in E)|$. Roughly, then, condition (1) asserts that X_n has a long-range tendency to lose track of its initial state.

Write $\mathscr{B}^{(n)} = \sigma[(\omega_{n'}); n' \geqq n]$, and recall that the tail σ-algebra $\mathscr{B}^{(\infty)}$ for a Markov chain is given by $\mathscr{B}^{(\infty)} := \bigcap_{n \in \mathbb{N}} \mathscr{B}^{(n)}$. We say that $\mathscr{B}^{(\infty)}$ is trivial if

$$P_i(B) = P_j(B) = 0 \textit{ or } 1 \qquad \text{for all } i, j \in S, \ B \in \mathscr{B}^{(\infty)}.$$

Using simple martingale arguments, one can show

Proposition 1. *Let X_n be a Markov chain. The following are equivalent:*

(i) *X_n has trivial tail field;*

(ii) *For any initial ν, and $A \in \mathscr{B}$,*

$$\lim_{n \to \infty} \sup_{B \in \mathscr{B}^{(n)}} |P_\nu(A \cap B) - P_\nu(A) P_\nu(B)| = 0;$$

(iii) *X_n is weakly ergodic;*

(iv) *All bounded f: $S \times \mathbb{N} \to \mathbb{R}$ such that*

$$f(i, n) = \sum_{k \in S} p_{ik} f(k, n+1)$$

are constant.

(Condition (ii) is called "mixing"; f as in (iv) are "space-time harmonic".) For rapid proof that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i), see [9]. These equivalences illustrate the usefulness of weak ergodicity as an expression of asymptotic loss of memory.

The next well-known result is due to Blackwell and Freedman (cf. [1] or [5]).

Theorem 1. *If X_n is an aperiodic random walk, or an aperiodic irreducible recurrent Markov chain, then $\mathscr{B}^{(\infty)}$ is trivial.*

Combining Proposition 1 and Theorem 1 one derives

Theorem 2. *Any aperiodic random walk is weakly ergodic.*

Theorem 3 (Orey). *Any aperiodic irreducible recurrent Markov chain is weakly ergodic.*

Remarks. The original proof [8] of Theorem 3 relied on an early version of Theorem 1, but not on (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) of Proposition 1; we will return to this point later in the discussion. Blackwell and Freedman [1] initiated a more detailed study of the atomic structure of $\mathscr{B}^{(\infty)}$ by partitioning S into so-called "cyclically moving subclasses". For i and j in the same class, $\lim_{n \to \infty} \delta_{ij}^{(n)} = 0$ if X_n is *any* random walk or recurrent chain.

We now introduce notation necessary for the definition of a Markov chain coupling. Let $\tilde{S}=S\times S$, and let $\tilde{X}_n=(X_n^1, X_n^2)$ be a bivariate stochastic process with coordinate representation $\tilde{\omega}=(\tilde{\omega}_0, \tilde{\omega}_1, \ldots)=((\omega_0^1, \omega_0^2), (\omega_1^1, \omega_1^2), \ldots)$ on $\tilde{\Omega}=\tilde{S}^{\mathbb{N}}$. It will sometimes be convenient to think of $\tilde{\Omega}$ as $S^{\mathbb{N}}\times S^{\mathbb{N}}$, and write $\tilde{\omega}=(\omega^1, \omega^2)$, where $\omega^1=(\omega_0^1, \omega_1^1, \ldots)$ and $\omega^2=(\omega_0^2, \omega_1^2, \ldots)$. Let $\tilde{P}_{(i,j)}$ be the measure on $(\tilde{\Omega}, \tilde{\mathscr{B}})$ for $\tilde{X}_n$ started in (i,j), with $\tilde{\mathscr{B}}=\sigma[(\tilde{\omega}_n);\ n\in\mathbb{N}]$. The *diagonal* D of $\tilde{S}$ is given by

$$D=\{(s,s)\colon\ s\in S\}.$$

For $\tilde{\omega}\in\tilde{\Omega}$, we define the hitting time to the diagonal, τ_D, by

$$\tau_D=\min\{n\in\mathbb{N}\colon\ \tilde{\omega}\in D\}$$
$$(=\infty \ \text{ if } \tilde{\omega}_n\notin D \ \text{ for all } n).$$

The following partition of $\tilde{\Omega}$ will be used throughout the discussion:

$$\begin{aligned}\tilde{\Omega}=\tilde{\Omega}^* &:=\{\tilde{\omega}\colon\ \tau_D(\tilde{\omega})<\infty,\ \text{ and } \tilde{\omega}_n\in D \text{ for all } n>\tau_D\}\\ +\tilde{\Omega}^\infty &:=\{\tilde{\omega}\colon\ \tau_D(\tilde{\omega})=\infty\}\\ +\tilde{\Omega}^0 &:=\tilde{\Omega}-(\tilde{\Omega}^*+\tilde{\Omega}^\infty).\end{aligned}$$

Let X_n be a Markov chain. In the context of the last paragraph, a *coupling for* X_n is a bivariate stochastic process $\tilde{X}_n$ such that for each $(i,j)\in\tilde{S}$,

$$\tilde{P}_{(i,j)}(\cdot\times\Omega)=P_i(\cdot) \quad\text{and}\quad \tilde{P}_{(i,j)}(\Omega\times\cdot)=P_j(\cdot), \tag{2a}$$

and

$$\tilde{P}_{(i,j)}(\tilde{\Omega}^*+\tilde{\Omega}^\infty)=1. \tag{2b}$$

Condition (2a) states that when $\tilde{X}_n$ starts in (i,j), the marginal processes X_n^1 and X_n^2 are copies of the given Markov chain X_n starting from i and j respectively; thus we may think of $\tilde{X}_n$ as two simultaneously evolving copies of X_n. Condition (2b) requires that $\tilde{X}_n$ remain on the diagonal D at all times after τ_D; we do not, however, rule out the possibility that τ_D might be infinite.

2. The Coupling Method

We now give a coupling criterion for weak ergodicity which has been known for many years in one form or another. The *existence* of useful couplings will be shown in later sections.

Proposition 2. *Let X_n be a Markov chain, and let $\tilde{X}_n$ be a coupling for X_n. If $\tilde{P}_{(i,j)}(\tau_D<\infty)=1$ for every $(i,j)\in\tilde{S}$, then X_n is weakly ergodic.*

Proof. By (2a) we have, for every $i,j,k\in S$,

$$\tilde{P}_{(i,j)}(\tilde{\omega}_n=(k,k))\leqq p_{ik}^{(n)}\wedge p_{jk}^{(n)}. \tag{3a}$$

Sum these inequalities over $k\in S$ to get

$$\tilde{P}_{(i,j)}(\tilde{\omega}_n\in D)\leqq\alpha_{ij}^{(n)}. \tag{3b}$$

But (2b) implies that $\{\omega_n\in D\}=\{\tau_D\leqq n\}$, so using the hypothesis of the proposition, it follows that

$$\liminf_{n\to\infty}\alpha_{ij}^{(n)}\geqq\lim_{n\to\infty}\tilde{P}_{(i,j)}(\tau_D\leqq n)=\tilde{P}_{(i,j)}(\tau_D<\infty)=1$$

for all $i,j\in S$. Since $\alpha_{ij}^{(n)}=1-\delta_{ij}^{(n)}$, we conclude that X_n is weakly ergodic.

In light of Proposition 2, we call a coupling $\tilde{X}_n$ such that $\tilde{P}_{(i,j)}(\tilde{\Omega}^*)=1$ for every $(i,j)\in\tilde{S}$ *successful*, and say that $\tilde{X}_n$ *fails* if $\tilde{P}_{(i,j)}(\tilde{\Omega}^*)<1$ for some $(i,j)\in\tilde{S}$.

3. Examples of Couplings

We now discuss briefly a few examples of couplings for Markov chains, and some of their applications.

Example 1. The classical coupling. Let $\tilde{X}_n$ be a Markov chain on $\tilde{S}$ with transition matrix $\tilde{p}=(\tilde{p}_{(i,j)(k,l)})_{(i,j),(k,l)\in\tilde{S}}$, given in terms of the matrix p for X_n by

$$\begin{array}{ccccl} \underline{\tilde{X}_n} & & \underline{\tilde{X}_{n+1}} & \underline{\tilde{p}_{..}} & \\ (i,i) & \rightarrow & (k,k) & p_{ik}; & (i,j,k,l\in S,\ i\neq j). \\ (i,j) & \rightarrow & (k,l) & p_{ik}\,p_{jl}; & \end{array}$$

It is a simple matter to check that $\tilde{p}$ determines a well-defined coupling. The process $\tilde{X}_n$ evolves as two *independent* copies of the chain X_n until these copies reach a common state; thereafter the marginal processes use the *same* transition mechanism. This was surely the first known coupling, dating back at least as far as a 1937 paper of Doeblin [4] – we will call it the *classical coupling*. Now suppose that X_n is irreducible and aperiodic. One can check that if X_n is positive rexurrent then $\tilde{X}_n$ is successful, while if X_n is transient then $\tilde{X}_n$ fails. In the null recurrent case, $\tilde{X}_n$ may either succeed or fail. An example of the latter type, described by Freedman [5, p. 45], with $S=\{0,1,\ldots\}$, has transition matrix p of the form

$$p_{0k}=c_k>0,\ p_{i\,i-1}=1; \qquad i,k\in S,\ i\neq 0, \tag{4}$$

for suitably chosen c_k. Thus, in general, the classical coupling yields weak ergodicity only for the positive recurrent case. In this setting Pitman [11] has noticed that $\tilde{E}_{(i,j)}(\tau_D)<\infty$; $(i,j)\in\tilde{S}$ (where $\tilde{E}$ is the expectation operator corresponding to $\tilde{P}$), which implies that we in fact have

$$\lim_{n\to\infty} n\,\delta_{ij}^{(n)}=0. \tag{5}$$

Pitman's paper contains a good deal of additional information relating to the classical coupling and its application to rate of convergence problems for positive recurrent chains.

Example 2. The Vasershtein coupling. For this example, let $\tilde{X}_n$ be a Markov chain on $\tilde{S}$ with $\tilde{p}$ of the form

$$\begin{array}{ccccl} \underline{\tilde{X}_n} & & \underline{\tilde{X}_{n+1}} & \underline{\tilde{p}_{..}} & \\ (i,i) & \rightarrow & (k,k) & p_{ik}\quad; & \\ (i,j) & \rightarrow & (k,k) & p_{ik}\wedge p_{jk}; & (i,j,k,l\in S,\ i\neq j,\ k\neq l) \\ (i,j) & \rightarrow & (k,l) & \dfrac{(p_{ik}-p_{jk})^+\,(p_{jl}-p_{il})^+}{1-\alpha_{ij}^{(1)}}. & \end{array}$$

It is not hard to verify that $\tilde{p}$ gives a coupling, the version for Markov chains of a more general construction due to Vasershtein [12]. Call this process $\tilde{X}_n$ the

Vasershtein coupling. Letting

$$\delta^{(n)} = \sup_{i,j} \delta_{ij}^{(n)},$$

we call a Markov chain *S-uniformly ergodic* if

$$\lim_{n\to\infty} \delta^{(n)} = 0.$$

The Vasershtein coupling turns out to be appropriate for the study of this property, providing results such as

Proposition 3 (Markov-Dobrushin-Hajnal). *Let X_n be a Markov chain. The following are equivalent:*

(i) *X_n is S-uniformly ergodic;*

(ii) *$\alpha^{(n)} = \inf_{i,j} \alpha_{ij}^{(n)} > 0$ for some $n \in \mathbb{N}$;*

(iii) *There are non-negative constants C and $\rho < 1$ such that $\delta^{(n)} \leqq C\rho^n$ for all $n \in \mathbb{N}$.*

For a coupling proof of Proposition 3, and a discussion of uniform ergodicity for nonstationary chains, see [6]. We note in passing that (iii) of Proposition 3 expresses geometric ergodicity, a more rapid rate of convergence than (5). Now the classical coupling is "weaker" than the Vasershtein coupling in the sense that under the former we go from (i,j) to $(k,k) \in D$ with probability $p_{ik} p_{jk}$, while under the latter the probability of this transition is $p_{ik} \wedge p_{jk}$, and $p_{ik} p_{jk} \leqq p_{ik} \wedge p_{jk}$ for all $i,j,k \in S$. Thus the second coupling may be successful for chains where the first fails, and so we might hope to get more general weak ergodicity results by using Vasershtein's process. Unfortunately the two couplings *agree* for chains of the form (4), since off the diagonal at least one of the two processes X_n^1 and X_n^2 is always moving deterministically. Thus, this coupling also fails for some null recurrent chains. Moreover, one can check that the present process sends us to the diagonal as efficiently as any *Markovian* coupling can, so that in order to prove Theorem 3 by coupling it is necessary to introduce a non-Markovian $\tilde{X}_n$.

Example 3. Ornstein's coupling. An ingenious coupling for random walk X_n on the integers $\mathbb{Z}$ has been given by Ornstein [10]; we will sketch his idea. Let $(p_k)_{k \in \mathbb{Z}}$ be the transition function for X_n (i.e. $p_{i\,i+k} \equiv p_k$; $i, k \in \mathbb{Z}$), and let N be a large positive integer to be chosen later. Roughly, to determine the progress of $\tilde{X}_n$ from time n to $n+1$, we observe the difference in the displacement of two independent copies of our random walk at this step. If the absolute value of this difference does not exceed N we let the copies move to their new states, while if the difference exceeds N we instead translate both processes by the *same* independently determined value k according to the distribution $\{p_k\}$. These heuristics lead to the *Ornstein coupling*, a Markov chain on $\tilde{S} = \mathbb{Z} \times \mathbb{Z}$ with transitions

$\tilde{X}_n$		$\tilde{X}_{n+1}$	$\tilde{p}_{..}$
(i,i)	$\to$	$(i+k, i+k)$	p_k;
(i,j)	$\to$	$(i+k, j+l)$	$p_k p_l \Delta_{\{\lvert k-l \rvert \leqq N\}} + [p_k - \Sigma_N(k)]\, \Delta_{\{k=l\}}$;

where

$$\Sigma_N(s) = \sum_{s': |s-s'| \leq N} p_s p_{s'}, \qquad \Delta_{\{\pi\}} = \begin{cases} 1 & \text{if } \pi \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

Until τ_D, $X_n^1 - X_n^2$ is a random walk with transition probabilities $q_{x\,x+y} \equiv q_y$ such that $q_y = q_{-y}$ and $q_y = 0$ for $y > N$. We therefore clearly have

$$\sum_{y \in \mathbb{Z}} |y| \, q_y < \infty \quad \text{and} \quad \sum_{y \in \mathbb{Z}} y q_y = 0.$$

By choosing N large enough we can also ensure that q is an aperiodic transition function, and for such an N, a classical random walk result by Chung and Fuchs [cf. 2, p. 270] states that q is recurrent. We conclude that $\tilde{P}_{(i,j)}(X_n^1 - X_n^2 = 0$ for some $n) = 1$, and hence the Ornstein coupling is successful. This is a second proof of Theorem 2. If we apply a variant on the above idea to the sojourn times of two copies of an irreducible aperiodic recurrent Markov chain from a reference state $s_0 \in S$, it is possible to construct a coupling for the chain which is always successful, yielding Theorem 3. Since this construction is based on the time space rather than the state space, the resulting process $\tilde{X}_n$ will be non-Markovian. Here, then, is one route to a coupling which succeeds in the null recurrent case, where Examples 1 and 2 fail. Another approach, of more general applicability, is the subject of the remainder of this paper.

4. A Maximal Coupling

In the proof of Proposition 2 we noted that (3 a) and (3 b) hold for an arbitrary coupling. Our objective now is to construct, for any given Markov chain X_n, a coupling $\tilde{X}_n$ such that these relations hold with equality, i.e. such that

$$\tilde{P}_{(i,j)}(\tilde{\omega}_n = (k,k)) = p_{ik}^{(n)} \wedge p_{jk}^{(n)}; \qquad i,j,k \in S, \ n \in \mathbb{N}, \tag{6}$$

and hence

$$\tilde{P}_{(i,j)}(\tau_D \leq n) = \alpha_{ij}^{(n)}.$$

For this coupling, it follows that

$$\tilde{P}_{(i,j)}(\tau_D < \infty) = 1 - \lim_{n \to \infty} \delta_{ij}^{(n)},$$

whence $\tilde{X}_n$ is successful if and only if X_n is weakly ergodic. The process $\tilde{X}_n$ will be called a *maximal coupling* for X_n; for the remainder of this paper $\tilde{X}_n$ will always denote such a coupling. The construction of $\tilde{X}_n$ is tedious, since the process is necessarily non-Markovian. Once we have shown its existence, though, the simple properties (2) and (6) can be used to advantage.

The main result of this paper is

Theorem 4. *Any Markov chain X_n has a maximal coupling $\tilde{X}_n$ satisfying* (6). *Thus $\tilde{X}_n$ is successful if and only if X_n is weakly ergodic.*

Proof. It suffices to prescribe consistent values of $\tilde{P}_{(i,j)}$; $(i,j) \in \tilde{S}$, for $\tilde{X}_n$ on cylinder sets of the form

$$\{\tilde{\omega}: \tilde{\omega}_1 = (i_1, j_1), \tilde{\omega}_2 = (i_2, j_2), \ldots, \tilde{\omega}_N = (i_N, j_N)\}; \qquad N \in \mathbb{N}, \tag{7}$$

and check that they satisfy (2a, b) and (7). The Kolmogorov extension theorem then guarantees a measure on $(\tilde{\Omega}, \tilde{\mathscr{B}})$ as desired. For brevity's sake we adopt the following notational conventions throughout the proof. The symbol i_n will be reserved for a generic value of ω_n^1, j_n for ω_n^2, and k_n for the case where $\omega_n^1 = \omega_n^2 = k_n$. We abbreviate

$$\begin{array}{rl} [s_n, s_{n'}] & \text{for } s_n, s_{n+1}, \ldots s_{n'}; \\ [(i_n, j_n), (i_{n'}, j_{n'})] & \text{for } (i_n, j_n), (i_{n+1}, j_{n+1}), \ldots (i_{n'}, j_{n'}); \\ [(k_n, k_n), (k_{n'}, k_{n'})] & \text{for } (k_n, k_n), (k_{n+1}, k_{n+1}), \ldots (k_{n'}, k_{n'}); \end{array}$$

all $i_n, j_n, k_n \in S$. Finally, any event of the form $\{\tilde{\omega}_{n_\xi} = (i_{n_\xi}, j_{n_\xi});\ \xi = 1, 2, \ldots, l\}$ will be written simply as $\{(i_{n_\xi}, j_{n_\xi});\ \xi = 1, 2, \ldots, l\}$ and similarly for events $\{\omega_{n_\xi}^1 = i_{n_\xi}\}$ and $\{\omega_{n_\xi}^2 = j_{n_\xi}\}$. Thus, for example, the $\tilde{P}_{(i,j)}$ measure of the event (7) will be denoted as

$$\tilde{P}_{(i,j)}([(i_1, j_1), (i_N, j_N)]).$$

We will now proceed to define $(\tilde{P}_{(i,j)})_{(i,j) \in \tilde{S}}$ in three steps. Parts (i) and (iii) are straightforward; part (ii), which contains the heart of the argument, requires some motivation. The idea is to construct $\tilde{P}_{(i,j)}$ so that for $N \geqq 1$,

$$\begin{aligned} &\tilde{P}_{(i,j)}([(i_1, j_1), (i_{N-1}, j_{N-1})]\,(k_N, k_N) \mid (k_N, k_N), \tau_D = N) \\ &\quad = \tilde{P}_{(i,j)}([i_1, i_{N-1}]\,k_N \mid \omega_N^1 = k_N, \tau_D \geqq N)\, \tilde{P}_{(i,j)}([j_1, j_{N-1}]\,k_N \mid \omega_N^2 = k_N, \tau_D \geqq N). \end{aligned}$$

Put $\varphi_{(i,j)}^{(N)}([s_1, s_N]) = \tilde{P}_{(i,j)}([s_1, s_N] \times \Omega \mid \omega_N^1 = s_N, \tau_D \geqq N)$, and

$$\psi_{(i,j)}^{(N)}(s) = \tilde{P}_{(i,j)}(\omega_N^1 = s, \tau_D \geqq N).$$

The equation above, together with the desired properties (2a, b) and (6), dictates the inductive prescription (9) for φ, equation (10) for ψ, and ultimately (14). However, a rigorous proof along these lines would be quite involved and not particularly enlightening, so we prefer to give the formal construction which follows.

(i) For $i = j$, set

$$\tilde{P}_{(i,j)}(\tilde{B}) = \begin{cases} P_i(B) = P_j(B) & \text{if } \tilde{B} = (B, B) \subset D^{\mathbb{N}} \\ 0 & \text{else} \end{cases},$$

$\tilde{B} \in \tilde{\mathscr{B}}$. Then $\tilde{P}_{(i,j)}(\tau_D = 0) = 1$, and (2a, b) and (6) are all obvious.

(ii) Henceforth, *assume* $i \neq j$, and define functions $\varphi_{(i,j)}^{(N)}\colon S^{\{1, \ldots, N\}} \to \mathbb{R}$ inductively by taking

$$\varphi_{(i,j)}^{(1)}(s_1) = 1 \qquad \text{for all } s_1 \in S, \tag{8}$$

and for $N \geqq 1$,

$$\varphi_{(i,j)}^{(N+1)}([s_1, s_{N+1}]) = \left(\frac{(p_{is_N}^{(N)} - p_{js_N}^{(N)})^+}{\psi_{(i,j)}^{(N+1)}(s_{N+1})} \right) p_{s_N s_{N+1}}\, \varphi_{(i,j)}^{(N)}([s_1, s_N]) \tag{9}$$

$$(= 0 \quad \text{if } \psi_{(i,j)}^{(N+1)}(s_{N+1}) = 0),$$

where

$$\psi_{(i,j)}^{(N)}(s) = \sum_{s' \in S} (p_{is'}^{(N-1)} - p_{js'}^{(N-1)})^+\, p_{s's}; \qquad s \in S,\ N \geqq 1. \tag{10}$$

These properties of the $\varphi^{(N)}_{(i,j)}$ are easily verified for $N \geqq 1$ by induction:

If

$$\varphi^{(N)}_{(i,j)}([s_1, s_N]) = 0, \text{ then } \varphi^{(N')}_{(i,j)}([s_1, s_{N'}]) = 0 \quad \text{for all } N' \geqq N. \tag{11}$$

$$\text{Either} \qquad \varphi^{(N+1)}_{(i,j)}([s_1, s_{N+1}]) = 0 \quad \text{or} \quad \varphi^{(N+1)}_{(j,i)}([s_1, s_{N+1}]) = 0 \tag{12}$$

(since either $(p^{(N)}_{is_N} - p^{(N)}_{js_N})^+$ or $(p^{(N)}_{js_N} - p^{(N)}_{is_N})^+$ is 0 in (9)).

$$\sum_{s_1, \ldots, s_{N-1} \in S} \varphi^{(N)}_{(i,j)}([s_1, s_N]) = \begin{cases} 1; & \psi^{(N)}_{(i,j)}(s_N) > 0 \\ 0; & \psi^{(N)}_{(i,j)}(s_N) = 0 \end{cases}. \tag{13}$$

We now specify the measures $\tilde{P}_{(i,j)}(\cdot \cap \tilde{\Omega}^*)$ on $(\tilde{\Omega}, \tilde{\mathscr{B}})$ for $(i,j) \in \tilde{S} - D$. To do so, first partition $\tilde{\Omega}^*$ as

$$\tilde{\Omega}^* = \sum_{N \in \mathbb{N}} \tilde{\Omega}^*_N, \qquad \text{with } \tilde{\Omega}^*_N = \{\tilde{\omega} \in \tilde{\Omega}^* \colon \tau_D(\tilde{\omega}) = N\}.$$

For each N, determine a measure on $\tilde{S}^{\{1, \ldots, N\}}$ by giving each of its atoms

$$\{[(i_1, j_1), (i_{N-1}, j_{N-1})] \, (k_N, k_N)\},$$

with $i_n \neq j_n$ *for* $1 \leqq n \leqq N-1$, the weight

$$\begin{aligned} &\tilde{P}_{(i,j)}([(i_1, j_1), (i_{N-1}, j_{N-1})](k_N, k_N)) \\ &\quad = [\psi^{(N)}_{(i,j)}(k_N) \wedge \psi^{(N)}_{(j,i)}(k_N)] \, \varphi^{(N)}_{(i,j)}([i_1, i_{N-1}] \, k_N) \, \varphi^{(N)}_{(j,i)}([j_1, j_{N-1}] \, k_N), \end{aligned} \tag{14}$$

and all other atoms weight 0. Now, for $N' > N$, let

$$\begin{aligned} &\tilde{P}_{(i,j)}([(i_1, j_1), (i_{N-1}, j_{N-1})] \, [(k_N, k_N), (k_{N'}, k_{N'})]) \\ &\quad = \tilde{P}_{(i,j)}([(i_1, j_1), (i_{N-1}, j_{N-1})] \, (k_N, k_N)) \prod_{\xi = N}^{N'-1} p_{k_\xi k_{\xi+1}}. \end{aligned} \tag{15}$$

These cylinder values are surely consistent, and so determine a measure

$$\tilde{P}_{(i,j)}(\cdot \cap \tilde{\Omega}^*_N)$$

concentrated on $\tilde{\Omega}^*_N$. Now, for $\tilde{B} \in \tilde{\mathscr{B}}$, let

$$\tilde{P}_{(i,j)}(\tilde{B} \cap \tilde{\Omega}^*) = \sum_{N \in \mathbb{N}} \tilde{P}_{(i,j)}(\tilde{B} \cap \tilde{\Omega}^*_N).$$

We claim that the following key properties are satisfied:

$$\tilde{P}_{(i,j)}(\tilde{\omega}_N = (k, k), \tilde{\omega} \in \tilde{\Omega}^*) = p^{(N)}_{ik} \wedge p^{(N)}_{jk}, \tag{16}$$

$$\tilde{P}_{(i,j)}([i_1, i_N], \tilde{\omega} \in \tilde{\Omega}^*) \leqq P_i([i_0, i_N]), \tag{17a}$$

$$= P_i([i_0, i_N]) \quad \text{if} \quad p^{(n)}_{ii_n} \leqq p^{(n)}_{ji_n} \quad \text{for some } n \leqq N, \tag{17b}$$

and similarly for $\tilde{P}_{(i,j)}([j_1, j_N], \tilde{\omega} \in \tilde{\Omega}^*)$.

To show (16), note that by (11) and (12), the right hand side of (14) is 0 if $i_n = j_n$ for some $n < N$. Thus we have

$$\begin{aligned} & P_{(i,j)}(\tilde\omega_N = (k,k), \tilde\omega \in \tilde\Omega_N^*) \\ &= \sum_{\substack{i_1, \ldots, i_{N-1} \in S \\ j_1, \ldots, j_{N-1} \in S}} [\psi_{(i,j)}^{(N)}(k) \wedge \psi_{(j,k)}^{(N)}(k)]\, \varphi_{(i,j)}^{(N)}([i_1, i_{N-1}]\, k)\, \varphi_{(j,i)}^{(N)}([j_1, j_{N-1}]\, k) \\ &= [\psi_{(i,j)}^{(N)}(k) \wedge \psi_{(j,i)}^{(N)}(k)], \end{aligned} \tag{18}$$

this last equality following from (13). Now for $N = 1$, (16) holds by (14), and a routine induction using (18) establishes (16) in general since

$$p_{ik}^{(N+1)} \wedge p_{jk}^{(N+1)} = [\psi_{(i,j)}^{(N+1)}(k) \wedge \psi_{(j,i)}^{(N+1)}(k)] - \sum_{s \in S} (p_{is}^{(N)} \wedge p_{js}^{(N)})\, p_{sk}.$$

The verification of (17) is slightly more involved. First, we show inductively that for all $N \geqq 1$,

$$\begin{aligned} & \tilde P_{(i,j)}([i_1, i_N], \tilde\omega \in \sum_{\xi=1}^{N} \tilde\Omega_\xi^*) \\ &\quad = P_i([i_1, i_N]) - (p_{i i_N}^{(N)} - p_{j i_N}^{(N)})^+\, \varphi_{(i,j)}^{(N)}([i_1, i_N]). \end{aligned} \tag{19}$$

For $N = 1$, this follows from (15). Assuming (19) for N, we have

$$\begin{aligned} & \tilde P_{(i,j)}\left([i_1, i_{N+1}], \tilde\omega \in \sum_{\xi=1}^{N+1} \tilde\Omega_\xi^*\right) \\ &= \tilde P_{(i,j)}\left([i_1, i_{N+1}], \tilde\omega \in \sum_{\xi=1}^{N} \tilde\Omega_\xi^*\right) + \tilde P_{(i,j)}([i_1, i_{N+1}], \tilde\omega \in \tilde\Omega_{N+1}^*) \\ &= [P_i([i_1, i_N]) - (p_{i i_N}^{(N)} - p_{j i_N}^{(N)})^+\, \varphi_{(i,j)}^{(N)}([i_1, i_N])]\, p_{i_N i_{N+1}} \\ &\quad + [\psi_{(i,j)}^{(N+1)}(i_{N+1}) \wedge \psi_{(j,i)}^{(N+1)}(i_{N+1})]\, \varphi_{(i,j)}^{(N+1)}([i_1, i_{N+1}]) \qquad \text{(by (13)–(15))} \\ &= P_i([i_1, i_{N+1}]) - \psi_{(i,j)}^{(N+1)}(i_{N+1})\, \varphi_{(i,j)}^{(N+1)}([i_1, i_{N+1}]) \qquad \text{(by (9))} \\ &\quad + [\psi_{(i,j)}^{(N+1)}(i_{N+1}) \wedge \psi_{(j,i)}^{(N+1)}(i_{N+1})]\, \varphi_{(i,j)}^{(N+1)}([i_1, i_{N+1}]) \\ &= P_i([i_1, i_{N+1}]) - (\psi_{(i,j)}^{(N+1)}(i_{N+1}) - \psi_{(j,i)}^{(N+1)}(i_{N+1}))^+\, \varphi_{(i,j)}^{(N+1)}([i_1, i_{N+1}]) \\ &= P_i([i_1, i_{N+1}]) - (p_{i i_{N+1}}^{(N+1)}) - (p_{j i_{N+1}}^{(N+1)})^+\, \varphi_{(i,j)}^{(N+1)}([i_1, i_{N+1}]), \end{aligned}$$

this last since $\psi_{(i,j)}^{(N+1)}(i_{N+1}) = p_{i i_{N+1}}^{(N+1)} - \sum_s (p_{is}^{(N)} \wedge p_{js}^{(N)})\, p_{s i_{N+1}}$.

We have proved (19) for all $N \geqq 1$, and hence

$$\tilde P_{(i,j)}\left([i_1, i_N], \tilde\omega \in \sum_{\xi=1}^{N} \tilde\Omega_\xi^*\right) \leqq P_i([i_1, i_N]) \qquad \text{for } N \geqq 1.$$

But for $N' > N$,

$$\begin{aligned} \tilde P_{(i,j)}\left([i_1, i_N], \tilde\omega \in \sum_{\xi=1}^{N'} \tilde\Omega_\xi^*\right) &= \sum_{i_{N+1}, \ldots, i_{N'} \in S} \tilde P_{(i,j)}\left([i_1, i_{N'}], \tilde\omega \in \sum_{\xi=1}^{N'} \tilde\Omega_\xi^*\right) \\ &\leqq \sum_{i_{N+1}, \ldots, i_{N'} \in S} P_i([i_1, i_{N'}]) = P_i([i_1, i_N]). \end{aligned}$$

Letting $N' \to \infty$, we obtain (20a). Moreover, if $p_{i i_n}^{(n)} \leqq p_{j i_n}^{(n)}$ for some $n < N$, then $\varphi_{(i,j)}^{(N)} = 0$ by (11), and (19) shows that

$$\tilde{P}_{(i,j)}\left([i_1, i_N], \tilde{\omega} \in \sum_{\xi=1}^{N} \tilde{\Omega}_{\xi}^{*}\right) = P_i([i_1, i_N]).$$

If $p_{i i_N}^{(N)} \leqq p_{j i_N}^{(N)}$, then this last equality again holds, also by (19). In either case,

$$\tilde{P}_{(i,j)}\left([i_1, i_N], \tilde{\omega} \in \sum_{\xi=N+1}^{\infty} \tilde{\Omega}_{\xi}^{*}\right) = 0$$

since $\varphi_{(i,j)}^{(\xi)}([i_1, i_\xi]) = 0$ for $\xi > N$ according to (9) and (11). Hence (17b) holds. The same argument clearly applies to the second marginal, so the verification of (17) is complete.

(iii) If $\tilde{P}_{(i,j)}(\tilde{\Omega}^*) = 1$, take $\tilde{P}_{(i,j)}(\tilde{\Omega}^\infty + \tilde{\Omega}^0) = 0$ to finish the construction of $\tilde{P}_{(i,j)}$. Then we must have equality in (17a), and similarly for the second marginal, so (2a) holds. Property (2b) is immediate, and (6) follows from (16). If $\tilde{P}_{(i,j)}(\Omega^*) = 1$ for every $(i, j) \in \tilde{S}$, then $\tilde{X}_n$ is a successful maximal coupling, and X_n is weakly ergodic.

(iv) If $\tilde{P}_{(i,j)}(\tilde{\Omega}^*) < 1$ for some $(i, j) \in \tilde{S}$, then using (17) we may define $\tilde{P}$ on cylinders $[(i_1, j_1)(i_N, j_N)]$; $i_n \neq j_n$ *for* $1 \leqq n \leqq N$, of $\tilde{\Omega}^\infty$ by

$$\tilde{P}_{(i,j)}([(i_1, j_1), (i_N, j_N)]) = \frac{[P_i([i_1, i_N]) - \tilde{P}_{(i,j)}([i_1, i_N], \tilde{\omega} \in \tilde{\Omega}^*)] \times [P_j([j_1, j_N]) - \tilde{P}_{(i,j)}([j_1, j_N], \tilde{\omega} \in \tilde{\Omega}^*)]}{1 - \tilde{P}_{(i,j)}(\Omega^*)}. \tag{20}$$

According to (17b), the right hand side of (20) vanishes if $i_n = j_n$ for some $n \leqq N$, so (23) extends to events $\tilde{B} \cap \tilde{\Omega}^\infty$ as

$$\tilde{P}_{(i,j)}\big((B^1, B^2) \cap \tilde{\Omega}^\infty\big) = \frac{[P_i(B_1) - \tilde{P}_{(i,j)}(B_1 \cap \tilde{\Omega}^*)]\,[P_j(B_2) - \tilde{P}_{(i,j)}(B_2 \cap \tilde{\Omega}^*)]}{1 - \tilde{P}_{(i,j)}\ (\tilde{\Omega}^*)};$$

$(B^1, B^2) \in \mathscr{B}$. Thus $\tilde{P}_{(i,j)}(\,\cdot \cap \tilde{\Omega}^\infty)$ is a product measure such that

$$\tilde{P}_{(i,j)}(\tilde{\Omega}^\infty) = 1 - \tilde{P}_{(i,j)}(\Omega^*),$$
$$\tilde{P}_{(i,j)}\big((B^1, \Omega)\big) = P_i(B^1) - \tilde{P}_{(i,j)}\big((B^1, \Omega) \cap \tilde{\Omega}^*\big),$$

and similarly for the second marginal. Setting $\tilde{P}_{(i,j)}(\tilde{\Omega}^0) = 0$, we obtain a coupling, and again, (6) is a consequence of (16). Hence $\tilde{X}_n$ is a maximal coupling, but in this case $\tilde{X}_n$ fails and X_n is not weakly ergodic. This completes the proof of the theorem.

*Remark.*The construction just completed extends easily to nonhomogeneous Markov chains, and more general discrete time Markov processes such as lattice interactions. For such processes we have solved a problem suggested by Vasershtein [12]:

If S is an abstract space, and P^1 and P^2 are two measures over $S^{\mathbb{N}}$, determine $\mu = (\mu_n)_{n \in \mathbb{N}}$, where

$$\mu_n = \inf_{\tilde{P}} \tilde{P}(\tilde{\omega}_n \in \tilde{S} - D),$$

$\tilde{P}$ a measure on $\tilde{S}^{\mathbb{N}}$ with projections P^1 and P^2. A more general study of this problem would be of interest.

5. An Application

In the last section we produced a maximal coupling $\tilde{X}_n$ for any given Markov chain X_n, which is absorbed on the diagonal of $\tilde{S}$ if and only if X_n is weakly ergodic. Unfortunately, the process $\tilde{X}_n$ is extremely complicated; it is probably too much to hope that its sample path behavior will be tractable in most cases. We can, however, make use of (2) and (6) to obtain

Theorem 5. *Let X_n be a Markov chain. X_n is weakly ergodic if and only if for every $i, j \in S$,*

$$P_i\left(p_{i\omega_n}^{(n)} \leqq p_{j\omega_n}^{(n)} \text{ for some } n \in \mathbb{N}\right) = 1$$

or (21)

$$P_j\left(p_{i\omega_n}^{(n)} \geqq p_{j\omega_n}^{(n)} \text{ for some } n \in \mathbb{N}\right) = 1$$

(or both).

Proof. If X_n is weakly ergodic, then $\mathscr{B}^{(\infty)}$ is trivial by Proposition 1. Let

$$B^1 = \{\omega \in \mathscr{B}: p_{i\omega_n}^{(n)} \leqq p_{j\omega_n}^{(n)} \text{ for infinitely many } n\},$$
$$B^2 = \{\omega \in \mathscr{B}: p_{i\omega_n}^{(n)} \geqq p_{j\omega_n}^{(n)} \text{ for infinitely many } n\}.$$

Then $B^1, B^2 \in \mathscr{B}^{(\infty)}$, with $B^1 \cup B^2 = \Omega$, so

$$P_i(B^1) = P_j(B^1) = 1 \quad or \quad P_i(B^2) = P_j(B^2) = 1.$$

Hence (21) holds. Conversely, suppose the first equality of (21) holds. Letting $\tilde{X}_n$ be the maximal coupling for X_n, we have

$$\tilde{P}_{(i,j)}\left(\tilde{\omega} \in \tilde{\mathscr{B}}: p_{i\omega_n^1}^{(n)} \leqq p_{j\omega_n^1}^{(n)} \text{ for some } n\right) = 1,$$

so, in particular, if we write $\tau^1(\tilde{\omega}) = \min\{n: p_{i\omega_n^1}^{(n)} \leqq p_{j\omega_n^1}^{(n)}\}$, then $\tilde{P}_{(i,j)}(\tau^1 < \infty) = 1$. But for each N, using (2) and (6) we have

$$\begin{aligned}\tilde{P}_{(i,j)}(\tau^1 = N, \tilde{\omega}_N \notin D) &\leqq \tilde{P}_{(i,j)}\left(p_{i\omega_N^1}^{(N)} \leqq p_{j\omega_N^1}^{(N)}, \tilde{\omega}_N \notin D\right) \\ &= \sum_{\substack{i_N \in S: \\ p_{ii_N}^{(N)} \leqq p_{ji_N}^{(N)}}} \left[\tilde{P}_{(i,j)}(\omega_N^1 = i_N) - \tilde{P}_{(i,j)}(\tilde{\omega}_N = (i_N, i_N))\right] \\ &= \sum_{\substack{i_N \in S: \\ p_{ii_N}^{(N)} \leqq p_{ji_N}^{(N)}}} (p_{ii_N}^{(N)} - p_{ji_N}^{(N)})^+ = 0.\end{aligned}$$

Thus $\tau_D \leqq \tau^1 < \infty$ a.s., so that $\tilde{X}_n$ is successful, and X_n is weakly ergodic. An analogous argument for the second possibility in (24) completes the proof.

It is interesting to note that once the existence of a maximal coupling is shown, Orey's original proof [8] of Theorem 3 may be interpreted as a coupling argument similar to the above. A proof along these lines is obtained by using Theorem 1 to claim that $\mathscr{B}^{(\infty)}$ is trivial, then considering the events B^1 and B^2 introduced during the proof of the last theorem, in order to deduce that $\tilde{X}_n$ is successful. In this way we avoid the notion of mixing ((ii) of Proposition 1) and a backward martingale argument required for the usual proof. Similar remarks apply to Theorem 2.

We conclude the discussion with two examples illustrating Theorem 5:

Example 1. $S=\mathbb{Z}$. X_n is a random walk such that

$$p_{i\,i+1}=p_{i\,i-1}=\frac{1}{4}, \quad p_{i\,i}=\frac{1}{2}; \quad i\in\mathbb{Z}.$$

Here $p_{ik}^{(n)}\leqq p_{jk}^{(n)}$ if and only if $|i-k|\geqq|j-k|$. Thus X_n is weakly ergodic if $P_i(|\omega_n-j|\leqq|\omega_n-i|$ for some $n)=1$. This is clearly the case, since $P_i(\omega_n=j$ for some $n)=1$.

Example 2. $S=\{0,1,\ldots\}$. X_n has transitions

$$p_{i\,0}=\varepsilon_i<1, \quad p_{i\,i+1}=1-\varepsilon_i; \quad i\in S,$$

with $\sum\varepsilon_i<\infty$. Consider $\bar{\omega}=(i,i+1,\ldots)$; by hypothesis $P_i(\bar{\omega})>0$, and for $j\neq i$, $p_{j\bar{\omega}_n}^{(n)}=p_{j\,i+n}^{(n)}=0<p_{i\bar{\omega}_n}^{(n)}$ for all n. Thus $P_i(p_{i\omega_n}^{(n)}\leqq p_{j\omega_n}^{(n)}$ for some $n)<1$. For the same i,j, $\bar{\bar{\omega}}=(j,j+1,\ldots)$ shows that also $P_j(p_{i\omega_n}^{(n)}\geqq p_{j\omega_n}^{(n)}$ for some $n)<1$, so X_n is not weakly ergodic. (In this example, it is also easy to see that $\bar{\omega},\bar{\bar{\omega}}\in\mathscr{B}^{(\infty)}$.)

Acknowledgements. Thanks to Professors F. Spitzer and R. Holley for their help.

References

1. Blackwell, D., Freedman, D.: The tail σ-field of a Markov chain and a theorem of Orey. Ann. Math. Statist. **35**, 1291–1295 (1964)
2. Chung, K. L.: A course in Probability Theory. 2nd ed. New York: Academic Press 1974
3. Dobrushin, R. L.: Markov processes with a large number of locally interacting components. Problemy Peredači. Informacii **7**, 70–87 (1971)
4. Doeblin, W.: Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états. Rev. Math. de l'Union Interbalkanique **2**, 77–105 (1937)
5. Freedman, D.: Markov Chains. San Francisco: Holden Day 1971
6. Griffeath, D.: Coupling methods for nonhomogeneous Markov chains. To appear
7. Harris, T.E.: Contact interactions on a lattice. Ann. Probab. **2**, 969–988 (1974)
8. Orey, S.: An ergodic theorem for Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete **1**, 174–176 (1962)
9. Orey, S.: Limit Theorems for Markov Chain Transition Probabilities. London: Van Nostrand 1971
10. Ornstein, D.: Random Walk I. T.A.M.S. **138**, 1–43 (1969)
11. Pitman, J. W.: Uniform rates of convergence for Markov chain transition probabilities. Z. Wahrscheinlichkeitstheorie verw. Gebiete **29**, 193–227 (1974)
12. Vasershtein, L. N.: Markov processes on countable product spaces describing large systems of automata. Problemy Peredači Informacii **3**, 64–72 (1969)

David Griffeath
Department of Mathematics
Cornell University
Ithaca, N.Y. 14850, USA

(Received November 8, 1974)