Z. Wahrscheinlichkeitstheorie verw. Gebiete
46, 193–204 (1979)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Maximal Coupling

Sheldon Goldstein*

Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

0. Introduction

Let $X_n^{(1)}$ and $X_n^{(2)}$ be (discrete time) stochastic processes. By a *coupling* of $X_n^{(1)}$ and $X_n^{(2)}$ we mean a simultaneous realization of these processes on the same probability space $(\Omega, \mathscr{B}, P)$. We say that a coupling is *successful* if the two processes eventually agree, i.e., if

$$P\{X_n^{(1)} = X_n^{(2)} \text{ for all } n \text{ sufficiently large}\} = 1.$$

Thus, a coupling of $X_n^{(1)}$ and $X_n^{(2)}$ may be regarded as a joint process $\tilde{X}_n = (\tilde{X}_n^{(1)}, \tilde{X}_n^{(2)})$ where $\tilde{X}_n^{(1)}$ is a copy of $X_n^{(1)}$ and $\tilde{X}_n^{(2)}$ is a copy on $X_n^{(2)}$; the coupling is successful if a.s. $\tilde{X}_n \in D$ for n sufficiently large, where D is the diagonal of the range space of $\tilde{X}_n$.

Successful coupling has been useful for proving ergodic theorems for Markov chains [1–5, 7]. In this context, $X_n^{(1)}$ and $X_n^{(2)}$ represent outcomes of the same Markov process beginning in two different states, and successful coupling is achieved by having the processes stay together once the same state (i.e., the diagonal) is reached. For a countable state space, the classical coupling, dating back at least to Doeblin [2], has the processes evolving independently until the diagonal is reached. This coupling is Markovian, but is inefficient because of the independence. A more efficient Markovian coupling has been used by Vaserstein [7]. Griffeath [3] has discovered a non-Markovian coupling which is maximal in the sense that the diagonal is attained as efficiently as possible (see below).

Coupling is also useful in ergodic theory: partitions characteristic of Bernoulli processes – called very weak Bernoulli partitions – can be defined using coupling [6].

In this paper we consider the following general question: When may two processes (not necessarily Markovian) be successfully coupled? We give necessary and sufficient conditions for this by constructing a maximal coupling for

* Supported in part by NSF Grant MPS 72-04534

0044-3719/79/0046/0193/$02.40

any two processes; this coupling will be successful if any successful coupling of the processes exists, and it brings the processes together as rapidly as possible. The result obtained also sheds some light on the significance of the tail σ-algebra.

1. Notation and Definitions

We will assume throughout that $X_n^{(i)}$, $i=1,2$, are stochastic processes taking values in the same standard Borel space $(\mathscr{E}, \mathscr{B})$.[1] We will also assume that $X_n^{(i)}$ are canonical realizations: Let $\Omega = \mathscr{E}^\infty = \overset{\infty}{\underset{1}{\times}} \mathscr{E}$ and $\mathscr{F} = \mathscr{B}^\infty = \overset{\infty}{\underset{1}{\times}} \mathscr{B}$. Let $P^{(i)}$, $i=1,2$, be two probability measures on $(\Omega, \mathscr{F})$. A point in Ω will be denoted by $\omega = (\omega_n)$, $\omega_n \in \mathscr{E}$. Let $X_n(\omega) = \omega_n$. Then the process $X_n^{(i)}$ is just X_n on $(\Omega, \mathscr{F}, P^{(i)})$.

Let $D = \{(x,x) | x \in \mathscr{E}\} \subset \mathscr{E} \times \mathscr{E}$ be the diagonal of $\mathscr{E} \times \mathscr{E}$. Since $(\mathscr{E}, \mathscr{B})$ is standard Borel, D is measurable in $(\mathscr{E} \times \mathscr{E}, \mathscr{B} \times \mathscr{B})$.

Let $\tilde{\Omega} = \Omega \times \Omega$. A point in $\tilde{\Omega}$ will be denoted by $\tilde{\omega} = (\omega^{(1)}, \omega^{(2)})$, $\omega^{(i)} \in \Omega$. We will identify $\tilde{\Omega}$ with $(\mathscr{E} \times \mathscr{E})^\infty$, so that we may write $\tilde{\omega} = (\tilde{\omega}_n)$, $\tilde{\omega}_n = (\omega_n^{(1)}, \omega_n^{(2)}) \in \mathscr{E} \times \mathscr{E}$. Let $\tilde{\mathscr{F}} = \mathscr{F} \times \mathscr{F}$, which may be identified with $(\mathscr{B} \times \mathscr{B})^\infty$.

Definition 1.1. A *coupling* of $P^{(1)}$ and $P^{(2)}$ is a probability measure $\tilde{P}$ on $(\tilde{\Omega}, \tilde{\mathscr{F}})$ with

$$\tilde{P}(\cdot \times \Omega) = P^{(1)} \quad \text{and} \quad \tilde{P}(\Omega \times \cdot) = P^{(2)}.$$

Definition 1.2. Let $\tilde{P}$ be a coupling of $P^{(1)}$ and $P^{(2)}$. $\tilde{P}$ is *successful* if

$$\tilde{P}\{\tilde{X}_n \in D \text{ for all } n \text{ sufficiently large}\} = 1.$$

Let $\mathscr{F}_n = \sigma\{X_m, m \geqq n\}$ be the smallest σ-algebra for which $X_n, X_{n+1}, \ldots$ are measurable. Let $\tilde{\mathscr{F}}_n = \mathscr{F}_n \times \mathscr{F}_n = \sigma\{\tilde{X}_m, m \geqq n\}$. Let $\mathscr{F}_\infty = \bigcap_{n=1}^{\infty} \mathscr{F}_n$ be the tail.

Let μ be a finite (signed) measure on the measurable space $\mathscr{X}$, let $\mu = \mu^+ - \mu^-$ be the Jordan decomposition of μ, and let $|\mu| = \mu^+ + \mu^-$. We denote by $\|\mu\| = |\mu|(\mathscr{X})$ the variation norm of μ. If ν is another finite measure on $\mathscr{X}$, the infimum $\mu \wedge \nu$ of μ and ν is given by

$$\mu \wedge \nu = 1/2(\mu + \nu - |\mu - \nu|). \tag{1}$$

$\mu \wedge \nu \geqq \lambda$, for any measure λ such that $\lambda \leqq \mu$ and $\lambda \leqq \nu$. It follows from (1) that if $\mu \geqq 0$ and $\nu \geqq 0$,

$$\|\mu \wedge \nu\| = 1/2(\|\mu + \nu\| - \|\mu - \nu\|).$$

In particular, if μ and ν are probability measures,

$$\|\mu \wedge \nu\| = 1 - 1/2\,\|\mu - \nu\|. \tag{2}$$

Let $P_n^{(i)} = P^{(i)} \restriction \mathscr{F}_n$ be the restriction of $P^{(i)}$ to $\mathscr{F}_n$ and let $P_\infty^{(i)} = P^{(i)} \restriction \mathscr{F}_\infty$. Let $\alpha_n = 1/2\,\|P_n^{(1)} - P_n^{(2)}\|$ and let $\alpha_\infty = \lim_{n \to \infty} \alpha_n$.

[1] For example, $\mathbb{R}^d$ or $\mathbb{Z}^d$

We denote by $\tilde{S}_n \subset \tilde{\Omega}$ the event "success occurs by time n." Thus, $\tilde{S}_n = \{\tilde{\omega} \in \tilde{\Omega} | \tilde{X}_m(\tilde{\omega}) \in D \text{ for } m \geqq n\}$. It is easy to see that for any coupling $\tilde{P}$,

$$\tilde{P}(\tilde{S}_n) \leqq 1 - \alpha_n. \tag{3}$$

(Let $\tilde{P}_{\tilde{S}_n} = \tilde{P}(\cdot \cap \tilde{S}_n)$. Then $\tilde{P}_{\tilde{S}_n} \leqq \tilde{P}$, so $\tilde{P}_{\tilde{S}_n}(\cdot \times \Omega) \leqq \tilde{P}(\cdot \times \Omega) = P^{(1)}$ and $\tilde{P}_{\tilde{S}_n}(\Omega \times \cdot) \leqq \tilde{P}(\Omega \times \cdot) = P^{(2)}$. But $\tilde{P}_{\tilde{S}_n}(\cdot \times \Omega) \upharpoonright \mathscr{F}_n = \tilde{P}_{\tilde{S}_n}(\Omega \times \cdot) \upharpoonright \mathscr{F}_n \equiv P_{\tilde{S}_n}$.[2] Thus $P_{\tilde{S}_n} \leqq P_n^{(1)}$ and $P_{\tilde{S}_n} \leqq P_n^{(2)}$. Therefore $P_{\tilde{S}_n} \leqq P_n^{(1)} \wedge P_n^{(2)}$. Thus $\tilde{P}(\tilde{S}_n) = \| P_{\tilde{S}_n} \| \leqq \| P_n^{(1)} \wedge P_n^{(2)} \| = 1 - \alpha_n$.)

Definition 1.3. A coupling $\tilde{P}$ is called *maximal* if

$$\tilde{P}(\tilde{S}_n) = 1 - \alpha_n. \tag{4}$$

2. Main Result

Theorem 2.1. *The following are equivalent:*

(i) *There exists a successful coupling $\tilde{P}$.*
(ii) $P_\infty^{(1)} = P_\infty^{(2)}$ *(agreement on tail)*
(iii) $\alpha_\infty = 0 (= \lim_{n \to \infty} \| (P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_n \|$.

Proof. We will prove (i) ⇒ (ii) ⇒ (iii) ⇒ (i). We will here prove (i) ⇒ (ii), which is easy. (iii) ⇒ (i) follows from the general existence of a maximal coupling, which will be constructed in the next section. (ii) ⇒ (iii) will be dealt with in Section 4.

We will actually prove that (i) ⇒ (iii) ⇒ (ii). (iii) ⇒ (ii) follows from the observation

$$2\alpha_n = \| (P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_n \| \geqq \| (P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_\infty \| = \| P_\infty^{(1)} - P_\infty^{(2)} \|. \tag{1}$$

That (i) ⇒ (iii) can be seen by noting that (i) $\Leftrightarrow \tilde{P}(\tilde{S}_n) \to 1$ as $n \to \infty \Rightarrow \alpha_n \to 0$ as $n \to \infty$ (by Eq. (3)).

Before proving (iii) ⇒ (i) we present a heuristic argument for (ii) ⇒ (i) containing the main ideas of the rigorous argument.

Imagine that $\mathscr{F}_\infty$ partitions Ω into fibers $\Omega(\omega)$. $\Omega(\omega)$ should be thought of as the collection of all $\omega' \in \Omega$ which end the same way as ω: $\omega' \in \Omega(\omega) \Leftrightarrow \omega'_n = \omega_n$ for n sufficiently large. Imagine also that the regular conditional probabilities P_ω for any probability measure P given $\mathscr{F}_\infty$ are concentrated on $\Omega(\omega)$: $P_\omega(\Omega(\omega)) = 1$. Then, if $P_\infty^{(1)} = P_\infty^{(2)} \equiv P_\infty$ we may define $\tilde{P}$ on $\mathscr{F}_\infty \times \mathscr{F}_\infty$ by lifting P_∞ to the diagonal of $\tilde{\Omega}$. $\tilde{P}$ on $\mathscr{F}_\infty \times \mathscr{F}_\infty$ could then be regarded as a measure on the set of fibers $(\Omega(\omega), \Omega(\omega'))$ (of $\tilde{\Omega}$ induced by $\mathscr{F}_\infty \times \mathscr{F}_\infty$) concentrated on the diagonal $- \{(\Omega(\omega), \Omega(\omega)) | \omega \in \Omega\}$. The extension $\tilde{P}$ to $\tilde{\mathscr{F}}$ could then be obtained by requiring that $\mathscr{F} \times \Omega$ and $\Omega \times \mathscr{F}$ be conditionally independent given $\mathscr{F}_\infty \times \mathscr{F}_\infty$. Thus on the fiber $(\Omega(\omega), \Omega(\omega))$ we would put the (conditional) probability distribution $P_\omega^{(1)} \times P_\omega^{(2)}$. For $\tilde{P}$ thus constructed, $\Omega(\omega^{(1)}) = \Omega(\omega^{(2)})$ for $\tilde{P}$ a.e. $\tilde{\omega} = (\omega^{(1)}, \omega^{(2)})$, i.e., $\omega^{(1)}$ and $\omega^{(2)}$ end the same way a.s.

The difficulty with the above argument stems from the fact that $\mathscr{F}_\infty$ is not (strictly) countably generated, so that, in particular, regular conditional proba-

2 This is the definition of $P_{\tilde{S}_n}$

bilities P_ω will not necessarily be concentrated on $\Omega(\omega)$. For example, if P has trivial tail, $P_\omega = P$ for all ω, but in general, P will not be concentrated on any $\Omega(\omega)$.

3. A Maximal Coupling

The maximal coupling $\tilde{P}$ will be constructed by piecing together a sequence of positive measures $\tilde{R}_n$ on $\tilde{\mathscr{F}}$. $\tilde{R}_n$ will be a measure for which "success begins at time n."

A basic element in the construction is the observation

$$(P_{n-1}^{(1)} \wedge P_{n-1}^{(2)}) \upharpoonright \mathscr{F}_n \leqq P_n^{(1)} \wedge P_n^{(2)}. \tag{1}$$

(Pf: $(P_{n-1}^{(1)} \wedge P_{n-1}^{(2)}) \upharpoonright \mathscr{F}_n \leqq P_{n-1}^{(1)} \upharpoonright \mathscr{F}_n = P_n^{(1)}$ and similarly, $(P_{n-1}^{(1)} \wedge P_{n-1}^{(2)}) \upharpoonright \mathscr{F}_n \leqq P_n^{(2)}$.) We thus define Q_n on $\mathscr{F}_n$ as follows:

$$Q_1 = P^{(1)} \wedge P^{(2)}$$

$$Q_n = (P_n^{(1)} \wedge P_n^{(2)}) - (P_{n-1}^{(1)} \wedge P_{n-1}^{(2)}) \upharpoonright \mathscr{F}_n \qquad \text{for } n > 1.$$

By (1), for all n

$$Q_n \geqq 0. \tag{2}$$

Q_n will be the main ingredient in $\tilde{R}_n$. We construct from Q_n a measure $\tilde{Q}_n$ on $(\tilde{\Omega}, \tilde{\mathscr{F}}_n)$ by "lifting Q_n to the diagonal": For $\tilde{A} = A^{(1)} \times A^{(2)} \in \tilde{\mathscr{F}}_n$,

$$\tilde{Q}_n(\tilde{A}) = Q_n(A^{(1)} \cap A^{(2)}). \tag{3}$$

Clearly,

$$\tilde{Q}_n \geqq 0, \tag{4}$$

and

$$\tilde{Q}_n(\tilde{\Omega} - S_n) = 0. \tag{5}$$

(This is perhaps most easily seen by noting that $\tilde{Q}_n$ is induced from Q_n by the map

$$\phi: \Omega \to \tilde{\Omega}$$

$$\omega \mapsto (\omega, \omega);$$

$\{\omega \in \Omega | \phi(\omega) \in \tilde{\Omega} - S_n\} = \phi$.)

To construct $\tilde{R}_n$ out of $\tilde{Q}_n$ we will use a sequence of kernels $K_n^{(i)}$: $\mathscr{F}_{n-1} \to \mathscr{F}_n$[3], $i = 1, 2$.

Definition 3.1. Let $(\mathscr{X}, \mathscr{H})$ be a standard Borel space and let $\mathscr{G} \subset \mathscr{H}$. Let $\lambda \geqq 0$ be a finite measure on $(\mathscr{X}, \mathscr{H})$. We denote by $K(\mathscr{H}, \mathscr{G}, \lambda) = K: \mathscr{X} \times \mathscr{H} \to [0, 1]$ the kernel giving the regular conditional probabilities for λ given $\mathscr{G}$[4]. This is defined by the following conditions:

---

3 $\mathscr{F}_k$ will denote the set of bounded $\mathscr{F}_k$ measurable functions as well as the σ-algebra

4 The fact that this is defined only λ a.e. presents no difficulty

(i) For $x \in \mathscr{X}$, $K(x, \cdot)$ is a probability measure on $\mathscr{H}$.
(ii) For $A \in \mathscr{H}$, $K(\cdot, A) \in \mathscr{G}$.
(iii) For $h \in \mathscr{H}$ and $g \in \mathscr{G}$

$$\int d\lambda\, g\, h = \int d\lambda\, g K h$$

where,

$$(K h)(x) = \int K(x, dy)\, h(y).$$

The following few statements refer to Definition 3.1: For any finite measure μ on $\mathscr{G}$, we denote by μK the measure on $\mathscr{H}$ given by

$$\int (d\mu K) f = \int d\mu (K f)$$

for any $f \in \mathscr{H}$.[5]

We note that

$$\|\mu K\| = \|\mu\|, \tag{6}$$

and that

$$\mu \geqq 0 \Rightarrow \mu K \geqq 0. \tag{7}$$

Furthermore,

$$g \in \mathscr{G} \Rightarrow K g = g \tag{8}$$

μ a.e., provided $\mu \ll \lambda \restriction \mathscr{G}$. In what follows, the measures μ with respect to which we would like (8) to hold a.e. will have the requisite absolute continuity. Moreover, it will be possible to choose a version of K for which (8) *strictly* holds. This is because of the product σ-algebra structure and the fact that the $\mathscr{F}_n$ are strictly countably generated. Thus, we might as well, and will, assume that the kernels K which we next define have this property.

We will denote $K(\mathscr{F}_n, \mathscr{F}_{n+1}, P_n^{(i)} - P_n^{(1)} \wedge P_n^{(2)})$ by $K_n^{(i)}: \mathscr{F}_n \to \mathscr{F}_{n+1}$, $i = 1, 2$. Let

$$\mathscr{K}_n^{(i)} = K_{n-1}^{(i)} K_{n-2}^{(i)} \dots K_2^{(i)} K_1^{(i)\,6}: \mathscr{F} \to \mathscr{F}_n \tag{9}$$

and let

$$\tilde{\mathscr{K}}_n = \mathscr{K}_n^{(1)} \times \mathscr{K}_n^{(2)}: \tilde{\mathscr{F}} \to \tilde{\mathscr{F}}_n. \tag{10}$$

$\tilde{\mathscr{K}}_n$ is defined by the condition

$$\tilde{\mathscr{K}}_n(f^{(1)} \times f^{(2)}) = \mathscr{K}_n^{(1)}(f^{(1)}) \times \mathscr{K}_n^{(2)}(f^{(2)}),$$

where $f^{(1)} \times f^{(2)}(\tilde{\omega}) = f^{(1)}(\omega^{(1)}) f^{(2)}(\omega^{(2)})$, $\quad f^{(i)} \in \mathscr{F}$.

We may now define $\tilde{R}_n$ by

$$\tilde{R}_n = \tilde{Q}_n \tilde{\mathscr{K}}_n \tag{11}$$

and $R_n^{(i)}$ by

$$R_n^{(i)} = Q_n \mathscr{K}_n^{(i)}. \tag{12}$$

---

[5] If $\mu \ll \lambda \restriction G$, then μK is independent of the choice of a version for K

[6] For $n = 1$, this product should be interpreted as the identity operator

$\tilde{R}_n$ is a finite measure on $\tilde{\mathscr{F}}$, while $\tilde{R}_n^{(i)}$ is a finite measure on $\mathscr{F}$, $i=1,2$. Furthermore, we easily have

Lemma 3.2.

(i) $\tilde{R}_n \geqq 0$

(ii) $R_n^{(i)} \geqq 0$

(iii) $\tilde{R}_n(\cdot \times \Omega) = R_n^{(1)}$

(iv) $\tilde{R}_n(\Omega \times \cdot) = R_n^{(2)}$

(v) $R_j^{(i)} \upharpoonright \mathscr{F}_k = Q_j \upharpoonright \mathscr{F}_k$ *for* $k \geqq j$.

(vi) $\sum_{j=1}^{m} R_j^{(i)} \upharpoonright \mathscr{F}_k = (P_m^{(1)} \wedge P_m^{(2)}) \upharpoonright \mathscr{F}_k$, *for* $k \geqq m$.

(vii) $\left\| \sum_{j=1}^{n} \tilde{R}_j \right\| = 1 - \alpha_n; \quad \left\| \sum_{j=1}^{\infty} \tilde{R}_j \right\| = \left\| \sum_{j=1}^{\infty} R_j^{(i)} \right\| = 1 - \alpha_\infty.$

Proof.

(i) follows from (11), (7), and (4).

(ii) follows from (12), (7), and (2).

(iii) and (iv) follow from (11), (10), and the fact that $\tilde{Q}_n(\cdot \times \Omega) = \tilde{Q}_n(\Omega \times \cdot) = Q_n$, which follows from (3).

(v) follows from (12) and (8).

(vi) follows from (v) and the computation

$$\sum_{j=1}^{m} Q_j \upharpoonright \mathscr{F}_k = (P^{(1)} \wedge P^{(2)}) \upharpoonright \mathscr{F}_k + [(P_2^{(1)} \wedge P_2^{(2)}) \upharpoonright \mathscr{F}_k - (P^{(1)} \wedge P^{(2)}) \upharpoonright \mathscr{F}_k] + \cdots$$
$$+ [(P_m^{(1)} \wedge P_m^{(2)}) \upharpoonright \mathscr{F}_k - (P_{m-1}^{(1)} \wedge P_{m-1}^{(2)}) \upharpoonright \mathscr{F}_k] = (P_m^{(1)} \wedge P_m^{(2)}) \upharpoonright \mathscr{F}_k.$$

(vii) follows from (vi) and (iii):

$$\left\| \sum_{j=1}^{n} \tilde{R}_j \right\| = \sum_{j=1}^{n} \tilde{R}_j(\tilde{\Omega}) = \sum_{j=1}^{n} R_j^{(i)}(\Omega)$$
$$= (P_n^{(1)} \wedge P_n^{(2)})(\Omega) = \| P_n^{(1)} \wedge P_n^{(2)} \| = 1 - \alpha_n.$$

The last equality follows from (1.2).

It follows from 3.2 vii that if $\alpha_\infty = 0$, $\sum_{j=1}^{\infty} \tilde{R}_j$ is a probability measure. For the general case, we proceed as follows:

Let $R_\infty^{(i)} = P^{(i)} - \sum_{j=1}^{\infty} R_j^{(i)}$, and let $\tilde{R}_\infty = \alpha_\infty^{-1} R_\infty^{(1)} \times R_\infty^{(2)}$ for $\alpha_\infty > 0$ and $\tilde{R}_\infty = 0$ for $\alpha_\infty = 0$. Then we define

$$\tilde{P} = \sum_{n=1}^{\infty} \tilde{R}_n + \tilde{R}_\infty. \tag{13}$$

Theorem 3.3. $\tilde{P}$, *defined in Equation* (13), *is a maximal coupling of* $P^{(1)}$ *and* $P^{(2)}$.

Proof. We must show several things: (a) that $\tilde{P}$ is a probability measure, (b) that $\tilde{P}$ has the right marginals, and (c) that $\tilde{P}$ is maximal. All of these follow from the next lemma.

Lemma 3.4. $\sum_{j=1}^{\infty} R_j^{(i)} \leqq P^{(i)}$.

Proof. It suffices to show that $\sum_{j=1}^{n} R_j^{(i)} \leqq P^{(i)}$ for all n.

By Lemma 3.2 vi,

$$\sum_{j=1}^{n} R_j^{(i)} \restriction \mathscr{F}_n = P_n^{(1)} \wedge P_n^{(2)} \leqq P_n^{(i)} = P^{(i)} \restriction \mathscr{F}_n.$$

Suppose that

$$\sum_{j=1}^{n} R_j^{(i)} \restriction \mathscr{F}_m \leqq P^{(i)} \restriction \mathscr{F}_m \qquad \text{for some } m \leqq n.$$

Then, we claim that

$$\sum_{j=1}^{n} R_j^{(i)} \restriction \mathscr{F}_{m-1} \leqq P^{(i)} \restriction \mathscr{F}_{m-1}; \tag{14}$$

by (reverse) induction, the lemma follows. To establish the claim (14) we observe that

$$\sum_{j=1}^{n} R_j^{(i)} \restriction \mathscr{F}_m \leqq P^{(i)} \restriction \mathscr{F}_m$$
$$\Rightarrow \sum_{j=m}^{n} R_j^{(i)} \restriction \mathscr{F}_m \leqq P_m^{(i)} - \sum_{j=1}^{m-1} R_j^{(i)} \restriction \mathscr{F}_m = (P^{(i)} - P_{m-1}^{(1)} \wedge P_{m-1}^{(2)}) \restriction \mathscr{F}_m \tag{15}$$

(from Lemma 3.2 vi).

It follows from the definition of the kernels $K_n^{(i)}$ that

$$[(P^{(i)} - P_{m-1}^{(1)} \wedge P_{m-1}^{(2)}) \restriction \mathscr{F}_m] \, K_{m-1}^{(i)} = (P^{(i)} - P_{m-1}^{(1)} \wedge P_{m-1}^{(2)}) \restriction \mathscr{F}_{m-1}$$
$$= P_{m-1}^{(i)} - P_{m-1}^{(1)} \wedge P_{m-1}^{(2)}.$$

It follows from (12) that

$$\left(\sum_{j=m}^{n} R_j^{(i)} \restriction \mathscr{F}_m \right) K_{m-1}^{(i)} = \sum_{j=m}^{n} R_j^{(i)} \restriction \mathscr{F}_{m-1},$$

since, for $m \leqq j$,

$$R_j^{(i)} \restriction \mathscr{F}_m = Q_j K_{j-1}^{(i)} \dots K_m^{(i)}.$$

Therefore, multiplying the LHS and the RHS of (15) on the right by $K_{m-1}^{(i)}$ we obtain (bearing in mind (7))

$$\sum_{j=m}^{n} R_j^{(i)} \restriction \mathscr{F}_{m-1} \leqq P_{m-1}^{(i)} - P_{m-1}^{(1)} \wedge P_{m-1}^{(2)} = P_{m-1}^{(i)} - \sum_{j=1}^{m-1} R_j^{(i)} \restriction \mathscr{F}_{m-1}$$

(from Lemma 3.2 vi).

Thus, it follows that

$$\sum_{j=1}^{n} R_j^{(i)} \upharpoonright \mathscr{F}_{m-1} \leqq P_{m-1}^{(i)} = P^{(i)} \upharpoonright \mathscr{F}^{m-1},$$

and the proof of the lemma is complete.

It follows from Lemma 3.4 that $R_\infty^{(i)} \geqq 0$. Thus $\tilde{P} \geqq 0$. To establish (a) and (b) consider separately the cases $\alpha_\infty = 0$ and $\alpha_\infty > 0$. First suppose $\alpha_\infty = 0$. Then by Lemma 3.2 vii,

$$\|\tilde{P}\| = \left\| \sum_{j=1}^{\infty} \tilde{R}_j \right\| = \left\| \sum_{j=1}^{\infty} R_j^{(i)} \right\| = 1.$$

This, together with Lemma 3.4, implies that

$$P^{(i)} = \sum_{j=1}^{\infty} R_j^{(i)} = \begin{cases} \tilde{P}(\cdot \times \Omega), & i=1 \\ \tilde{P}(\Omega \times \cdot), & i=2. \end{cases}$$

Suppose next that $\alpha_\infty > 0$. Then, since by Lemma 3.4,

$$\|R_\infty^{(i)}\| = \|P^{(i)}\| - \left\| \sum_{j=1}^{\infty} R_j^{(i)} \right\| = 1 - (1 - \alpha_\infty) = \alpha_\infty,$$

we have

$$\begin{aligned} \tilde{P}(\cdot \times \Omega) &= \sum_{n=1}^{\infty} \tilde{R}_n(\cdot \times \Omega) + \tilde{R}_\infty(\cdot \times \Omega) \\ &= \sum_{n=1}^{\infty} R_n^{(1)} + \alpha_\infty^{-1} R_\infty^{(1)}(\cdot) R_\infty^{(2)}(\Omega) \\ &= \sum_{n=1}^{\infty} R_n^{(1)} + \alpha_\infty^{-1} \left(P^{(1)} - \sum_{n=1}^{\infty} R_n^{(1)} \right) \alpha_\infty \\ &= P^{(1)}, \end{aligned}$$

and similarly,

$$\tilde{P}(\Omega \times \cdot) = P^{(2)}.$$

This establishes that $\tilde{P}$ is a probability measure with the right marginals.

$\tilde{P}$ is maximal because,

$$\begin{aligned} \tilde{P}(\tilde{S}_n) &\geqq \sum_{j=1}^{n} \tilde{R}_j(\tilde{S}_n) = \sum_{j=1}^{n} \tilde{Q}_j(\tilde{S}_n) \\ &= \sum_{j=1}^{n} \tilde{Q}_j(\tilde{\Omega}) = \sum_{j=1}^{n} \|\tilde{Q}_j\| = \sum_{j=1}^{n} \|\tilde{R}_j\| = \left\| \sum_{j=1}^{n} \tilde{R}_j \right\| = 1 - \alpha_n. \end{aligned} \tag{16}$$

Here, we have used (8), (11), (5), (6), and Lemma 3.2 vii, as well as the fact that $\tilde{S}_n \in \tilde{F}_j$ for $j \leqq n$. In particular, the step $\tilde{R}_j(\tilde{S}_n) = \tilde{Q}_j(\tilde{S}_n)$ follows from the fact that

$$\tilde{R}_j \upharpoonright \tilde{F}_j = \tilde{Q}_j \mathscr{K}_j \upharpoonright \tilde{\mathscr{F}}_j = \tilde{Q}_j \upharpoonright \tilde{\mathscr{F}}_j,$$

which need only be checked for product functions: For $f^{(1)}\in\mathscr{F}_j$ and $f^{(2)}\in\mathscr{F}_j$,

$$\begin{aligned}\int d\tilde{Q}_j\,\tilde{\mathscr{K}}_j(f^{(1)}\times f^{(2)})&=\int d\tilde{Q}_j(\mathscr{K}_j^{(1)}(f^{(1)})\times\mathscr{K}_j^{(2)}(f^{(2)}))\\&=\int d\tilde{Q}_j(f^{(1)}\times f^{(2)})\qquad (=\int dQ_j f^{(1)}f^{(2)}).\end{aligned}$$

This completes the proof of Theorem 3.3.

Remark 3.5. The maximal coupling $\tilde{P}$ depends only upon $P^{(1)}$ and $P^{(2)}$, not upon the versions of the kernels $K_n^{(i)}$. This may easily be seen by applying $\tilde{R}_n$ to product functions, using the fact, refering to Definition 3.1, that K is well defined and version independent on functions mod λ (i.e., as a map from $\mathscr{H}(\text{mod}\,\lambda)$ to $\mathscr{G}(\text{mod}\,\lambda)$).

Remark 3.6. Let $\tilde{S}$ be the event "success occurs." Then

$$\tilde{S}=\bigcup_n \tilde{S}_n,$$

and $\sum_{j=1}^{\infty}\tilde{R}_j$ is "supported" by $\tilde{S}$:

$$\sum_{j=1}^{\infty}\tilde{R}_j(\tilde{\Omega}-\tilde{S})\leqq\sum_{j=1}^{\infty}\tilde{R}_j(\tilde{\Omega}-\tilde{S}_j)=0.$$

Moreover, $\tilde{R}_\infty$ is "supported" by $\tilde{\Omega}-\tilde{S}$: Since $\tilde{P}(\tilde{S}_n)=1-\alpha_n=\sum_{j=1}^{\infty}\tilde{R}_j(\tilde{S}_n)$ (by Eq. (16)), we have that $\tilde{R}_\infty(\tilde{S})=\lim_{n\to\infty}\tilde{R}_\infty(\tilde{S}_n)=0$.

Remark 3.7. Suppose $\tilde{P}$ is any maximal coupling. Then

$$\tilde{P}(\tilde{S})=\lim_{n\to\infty}\tilde{P}(\tilde{S}_n)=\lim_{n\to\infty}(1-\alpha_n)=1-\alpha_\infty.$$

In particular, if $\alpha_\infty=0$, $\tilde{P}(\tilde{S})=1$ and $\tilde{P}$ is successful. Thus (iii) $\Rightarrow$ (i) of Theorem 2.1 follows from Theorem 3.3.

4. Tail Agreement

We consider the implication (ii) $\Rightarrow$ (iii) of Theorem 2.1. For any n, there exists a set $A_n\in\mathscr{F}_n$ such that

$$|P^{(1)}(A_n)-P^{(2)}(A_n)|\geqq\alpha_\infty.$$

Without loss of generality we may, by passing to a subsequence, assume that

$$P^{(1)}(A_n)-P^{(2)}(A_n)\geqq\alpha_\infty \tag{1}$$

for all n.

Let $\mathfrak{H}=L^2(\Omega,\mathscr{F},P^{(1)}+P^{(2)})$.

Since the sequence $X_n'\equiv I_{A_n}$ (I_A is the indicator function of the set A: $I_A(\omega)=1$ if $\omega\in A$ and $=0$ otherwise) is norm-bounded in $\mathfrak{H}$, it has a weakly convergent subsequence

$$X'_{n_j} \xrightarrow{w} X'_\infty.$$

Let $E^{(i)}(f) = \int f dP^{(i)}$ for $f \in \mathfrak{H}$. Since $|E^{(i)}(f)| \leqq E^{(i)}(|f|) \leqq \int |f| d(P^{(1)} + P^{(2)}) \leqq \sqrt{2} \|f\|_2$, $E^{(i)} \in \mathfrak{H}^*$. Thus,

$$E^{(1)}(X'_\infty) - E^{(2)}(X'_\infty) = \lim_{j\to\infty} (P^{(1)}(A_{n_j}) - P^{(2)}(A_{n_j})) \geqq \alpha_\infty.$$

Since subspaces are weakly closed, $X'_\infty \in \mathscr{F}_\infty$, and we are done.[7]

In fact, we have that $0 \leqq X'_\infty \leqq 1$ so that

$$\tilde{\alpha}_\infty \equiv 1/2 \|(P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_\infty\| \geqq \alpha_\infty.$$

Since, by Equation (2.1), $\alpha_\infty \geqq \tilde{\alpha}_\infty$, we have proven the following:

Proposition 4.1. $\|(P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_\infty\| = \lim_{n\to\infty} \|(P^{(1)} - P^{(2)}) \upharpoonright \mathscr{F}_n\|$.

5. Additional Remarks

(1) Let $\tilde{T}_n$ be the event "success begins at time n," i.e.,

$$\tilde{T}_n = \tilde{S}_n - \tilde{S}_{n-1}.$$

Let $\tilde{P} = \sum_{n=1}^{\infty} \tilde{R}_n + \tilde{R}_\infty$ be as in Section 3. Then $\tilde{R}_n$ is "supported" by $\tilde{T}_n$ (since, by (16),

$$1 - \alpha_{n-1} = \sum_{j=1}^{n-1} \tilde{R}_j(\tilde{S}_{n-1}) \leqq \sum_{j=1}^{n} \tilde{R}_j(\tilde{S}_{n-1}) \leqq \tilde{P}(\tilde{S}_{n-1}) = 1 - \alpha_{n-1},$$

so that $\tilde{R}_n(\tilde{S}_{n-1}) = 0$).

(2) Suppose $P^{(1)}$ and $P^{(2)}$ represent the same Markov chain (with countable state space) starting in different states, say 1 and 2. Then the maximal coupling $\tilde{P}$ constructed in Section 3 has the following 2 properties: Let τ_D be the hitting time of the diagonal:

$$\tau_D = \min\{n | \tilde{X}_n \in D\}.$$

(If $\tilde{X}_n \in \tilde{\Omega} - D$ for all n, τ_D is defined to be ∞.) Then,

(a) $\tilde{S}_n = \{\tau_D \leqq n\}$ (mod $\tilde{P}$), i.e.,

$$\tilde{X}_n \in D \quad \text{for } n \geqq \tau_D \, \tilde{P} \quad \text{almost surely,}$$

and (b) $\omega_n^{(1)}$, $n \leqq \tau_D$ and $\omega_n^{(2)}, n \leqq \tau_D$ are conditionally independent given τ_D and $\tilde{X}_{\tau_D}$. Also, (c) $\|P_n^{(1)} - P_n^{(2)}\| = \|P_{1,\cdot}^n - P_{2,\cdot}^n\|$, where P is the transition matrix. ((c)

[7] I am grateful to S. Varadhan for bringing this argument to my attention

follows easily from the Markov property. (a) follows from (c) and the maximality of $\tilde{P}$, and (b) follows from (a), remark (1), and the Markov property). In fact, $\tilde{P}$ in this case turns out to be precisely the coupling of Griffeath [3], a result strongly suggested by the proposition of the next remark.

(3) The coupling $\tilde{P}$ has the property that

(I) $\mathscr{F} \times \Omega$ and $\Omega \times \mathscr{F}$ are conditionally independent on $\tilde{\Omega} - \tilde{S}_{n-1}$ given $\mathscr{F}_n \times \Omega$, and given $\Omega \times \mathscr{F}_n$; and $\mathscr{F} \times \Omega$ and $\Omega \times \mathscr{F}$ are independent on $\tilde{\Omega} - \tilde{S}$. In fact, for $A \in \mathscr{F}$

$$\tilde{P}\{A \times \Omega | \tilde{\Omega} - \tilde{S}_{n-1}; \omega_m^{(1)}, m \geqq n; \omega_m^{(2)}, m \geqq 1\}$$
$$= \mathscr{K}_n^{(1)}(\omega^{(1)}, A) \quad \text{for } \tilde{P} \quad \text{a.e.} \quad \tilde{\omega} = (\omega^{(1)}, \omega^{(2)}).$$

We will say that a coupling satisfying (I) is conditionally independent. It is not difficult to see that $\tilde{P}$ is characterized by conditional independence and maximality:

Proposition. *$\tilde{P}$ is the unique maximal conditionally independent coupling of $P^{(1)}$ and $P^{(2)}$.*

(4) Griffeath [3] raises the following question: If $\mathscr{E}$ is an abstract space, and $P^{(1)}$ and $P^{(2)}$ are two measures on $\Omega = \mathscr{E}^\infty$, determine $\mu = (\mu_n)_{n \in \mathbb{N}}$, where

$$\mu_n = \inf_{\tilde{P}} \tilde{P}\{\tilde{\omega}_n \in \tilde{\Omega} - D\},$$

$\tilde{P}$ a coupling of $P^{(1)}$ and $P^{(2)}$.

It is easy to see, at least if $\mathscr{E}$ is a standard Borel space, by using the version of (2.3) appropriate to the present context, that

$$\mu_n = 1/2 \|(P^{(1)} - P^{(2)}) \upharpoonright \mathscr{B}_n\|,$$

where $\mathscr{B}_n = \sigma\{X_n\} \subset \mathscr{F}$, and that, in fact, μ_n is realized for some coupling $\tilde{P}$, which may be constructed using the method of Section 3.

(5) Theorem 2.1 provides a precise formulation of the notion that tail information tells us how a process ends, that two processes agree on the tail if and only if they "end the same way."

(6) Suppose $P_\infty^{(1)} = P_\infty^{(2)}$. It is perhaps worthwhile to indicate what goes wrong if one tries to construct a successful coupling beginning with the tail: Let $Q_\infty = P_\infty^{(1)} \wedge P_\infty^{(2)} = P_\infty^{(1)} = P_\infty^{(2)}$. Lifting Q_∞ to the diagonal we obtain $\tilde{Q}_\infty$ on $\mathscr{F}_\infty \times \mathscr{F}_\infty$. Let $\mathscr{K}_\infty^{(i)} = \mathscr{K}(\mathscr{F}, \mathscr{F}_\infty, \mathrm{P}^{(i)})$, and let $\tilde{\mathscr{K}}_\infty = \mathscr{K}_\infty^{(1)} \times \mathscr{K}_\infty^{(2)}$. Finally, let $\tilde{P} = \tilde{Q}_\infty \tilde{\mathscr{K}}_\infty$.

The difficulty is that $\mathscr{F}_\infty \times \mathscr{F}_\infty \neq \bigcap_n \tilde{\mathscr{F}}_n$; in fact, $\tilde{S} \notin \mathscr{F}_\infty \times \mathscr{F}_\infty$. Thus, it is not true that $\tilde{Q}_\infty(\tilde{S}) = 1$. (If $\tilde{S}$ were $\mathscr{F}_\infty \times \mathscr{F}_\infty$ measurable, it would follow, just as for the corresponding result for $\tilde{Q}_n$, that $\tilde{Q}_\infty(\tilde{S}) = 1$.)

Acknowledgements. I wish to thank Steve Kalikow for raising questions which led to the problem with which this paper is concerned. I also wish to thank David Griffeath for bringing to my attention the connection between the present work and his own work.

References

1. Dobrushin, R.L.: Markov processes with a large number of locally interacting components. Problemy Peredači Informacii **7**, 70–87 (1971)
2. Doeblin, W.: Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'etats. Rev. Math. de l'Union Interbalkanique **2**, 77–105 (1937)
3. Griffeath, D.: A maximal coupling for Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 95–106 (1975)
4. Griffeath, D.: Uniform coupling of non-homogeneous Markov chains. J. Appl. Probability **12**, 753–762 (1975)
5. Harris, T.E.: Contact interactions on a lattice. Ann. Probability **2**, 969–988 (1974)
6. Ornstein, D.: Ergodic theory, randomness, and dynamical systems. New Haven and London: Yale Univ. Press 1974
7. Vasershtein, L.N.: Markov processes on countable product spaces describing large systems of automata. Problemy Peredači Informacii **3**, 64–72 (1969)

Received April 1, 1978