Z. Wahrscheinlichkeitstheorie verw. Gebiete
35, 315–322 (1976)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On Coupling of Markov Chains*

J. W. Pitman

Statistical Laboratory, Dept. of Statistics, University of California, Berkeley, Cal. 94720, USA

1. Introduction

Let $(X_n, n \in N)$ be a homogeneous Markov chain with discrete time set $N = \{0, 1, 2, \ldots\}$ and countable state space S. Let μ be the initial distribution, p the transition matrix, and say that (X_n) is Markov (μ, p). Let μ_n denote the distribution μp^n of X_n,

$$\mu_n(y) = \mu p^n(y) = \sum_{x \in S} \mu(x) p^n(x, y), \qquad y \in S, \ n \in N,$$

and consider the behavior of μ_n as $n \to \infty$. It may be that there is a probability π on S such that $\mu_n(y) \to \pi(y)$ as $n \to \infty$ for all $y \in S$. In this case π is necessarily p-invariant and μ_n converges uniformly to π over all subsets of S:

$$\lim_{n \to \infty} \|\mu_n - \pi\| = 0, \tag{1}$$

where for two probabilities λ and λ' on S

$$\|\lambda - \lambda'\| = \sup_{A \in S} |\lambda(A) - \lambda'(A)| = \tfrac{1}{2} \sum_{y \in S} |\lambda(y) - \lambda'(y)|$$

denotes half the usual total variation distance between λ and λ'. More generally, there may be no p-invariant probability π so that (1) is impossible, but still for two given initial distributions μ and μ' the distributions $\mu_n = \mu p^n$ and $\mu'_n = \mu' p^n$ may become arbitrarily close as $n \to \infty$ in the sense that

$$\lim_{n \to \infty} \|\mu_n - \mu'_n\| = 0. \tag{2}$$

Here (1) is the special case of (2) with $\mu' = \pi$, and (2) holds for all μ and μ' if (1) holds for all μ with the same π. Note that in any case the sequences of norms appearing in (1) and (2) are monotone decreasing since p acts as a contraction on bounded signed measures with total variation norm. For an irreducible, aperiodic,

* This research was prepared with the support of National Science Foundation grant MPS-75-10376

and recurrent Markov chain it is known that (2) always holds, with (1) holding iff the chain is positive recurrent (see Freedman [2]). Similar behavior is encountered when this kind of convergence in total variation norm is considered for φ-recurrent chains with general measurable state space (see Orey [10]).

There is an elegant general method for establishing convergence and rates of convergence in (2) which is due to Doeblin [1]. This coupling method relies on the simple observation that if two probabilities λ and λ' on S are the distributions of S valued random variables X and X' defined on the same probability space $(\Omega, \mathscr{F}, P)$, then

$$\|\lambda-\lambda'\| \leqq P(X \neq X'). \tag{3}$$

Thus if on some probability space $(\Omega, \mathscr{F}, P)$ we can define a Markov (μ, p) chain (X_n) and a Markov (μ', p) chain (X'_n) so that $X_n(\omega)=X'_n(\omega)$ for all $n \geqq T(\omega)$, where $\omega \in \Omega$ and $T(\omega)$ is a random time, then $(X_n \neq X'_n) \subset (T>n)$, whence

$$\|\mu_n-\mu'_n\| \leqq P(T>n). \tag{4}$$

Following Griffeath [4] we shall say that such a set up provides a coupling of a Markov (μ, p) chain and a Markov (μ', p) chain, though note that we do not require as Griffeath does that T should be the first time that the two chains meet. If $P(T<\infty)=1$ the coupling is said to be *successful*, and we deduce at once that (2) holds. Again, if T has finite expectation the convergence is $o(1/n)$ and the sequence of norms has finite sum, and correspondingly faster rates of convergence are obtained if T possesses higher moments. For some effective applications of this method see Griffeath [4–7] and Pitman [11]. References may be found in [4] to the use of coupling arguments in the theory of Markovian lattice interactions, and for a different application of the basic bound (3) to Poisson convergence see Hodges and LeCam [8], Freedman [3], and Serfling [13].

The main result of Griffeath's paper [4] is that existence of a successful coupling is not only sufficient but necessary for the convergence (2). Indeed, Griffeath establishes the existence of a *maximal coupling* which attains equality in the inequality (4). The main object of the present paper is to provide a new and much simplified construction of Griffeath's maximal coupling.

The prototype for this construction is the following simple scheme for attaining equality in (3): given two probabilities λ and λ' on S, let the joint distribution of X and X' be specified as follows:

$$P(X=X'=x)=\lambda(x) \wedge \lambda'(x), \qquad x \in S,$$
$$P(X=x, X'=x')=[(\lambda-\lambda')^+(x)]\,[(\lambda-\lambda')^-(x')]/\|\lambda-\lambda'\|, \quad x, x' \in S, \quad x \neq x'.$$

Thus X and X' are made to agree with as great a probability as possible, and then given that they differ they are made conditionally independent with the mutually singular conditional distributions $(\lambda-\lambda')^+/\|\lambda-\lambda'\|$ and $(\lambda-\lambda')^-/\|\lambda-\lambda'\|$ which are required to make X have law λ and X' law λ'. For some instances of this maximal coupling of two random variables see Vasershtein [14] and Serfling [13].

For the maximal coupling of two Markov chains it is necessary to provide simultaneously for all n a maximal coupling of random variables X_n and X'_n which keeps the sequences (X_n) and (X'_n) evolving as Markov chains. This is achieved by a *path decomposition* in the spirit of Williams [16] (see also [9]). The required joint law of two Markov chains (X_n) and (X'_n), which agree after the random time T when they first meet, is specified by first giving the distribution in space-time of the junction (T, X_T), then the conditional law given T and X_T of the two pre-T processes and the single post-T process, with the requirement that these three fragments be conditionally independent given T and X_T.

To formulate this result let $\tilde{\Omega}$ be the space of all sequences

$$\tilde{\omega}=((\omega_0, \omega'_0), (\omega_1, \omega'_1), \dots)$$

of pairs of points in S, and equip $\tilde{\Omega}$ with the product σ-field $\tilde{\mathscr{F}}$ generated be the coordinate maps $X_0, X_1, \dots, X'_0, X'_1, \dots$, where, for instance, $X_n(\tilde{\omega})=\omega_n$. Let $T(\tilde{\omega})$ be the first n such that $X_n(\tilde{\omega})=X'_n(\tilde{\omega})$, with $T(\tilde{\omega})=\infty$ if there is no such n.

Theorem. *Let μ and μ' be two mutually singular probabilities on S, let $\alpha_n=\mu p^n-\mu' p^n$, and suppose that $\lim_{n\to\infty}\|\alpha_n\|=0$. There exists a unique probability $\tilde{P}$ on $(\tilde{\Omega}, \tilde{\mathscr{F}})$ such that*

(i) $\tilde{P}(T=n, X_n=z)=(\alpha^+_{n-1}\,p-\alpha^+_n)(z)$, *$n\geqq 1$, $z\in S$, and*

(ii) *under $\tilde{P}$ conditional on $(T=n, X_n=z)$, for each $n\geqq 1$, $z\in S$,*

(a) *the two pre-T processes $(X_0, \dots, X_n)$ and $(X'_0, \dots, X'_n)$ are inhomogeneous Markov chains with reverse transition probabilities from y at time m to x at time $m-1$ given by $\alpha^+_{m-1}(x)\,p(x,y)/\alpha^+_{m-1}\,p(y)$ and $\alpha^-_{m-1}(x)\,p(x,y)/\alpha^-_{m-1}\,p(y)$, respectively,*

(b) *the post-T processes $(X_n, X_{n+1}, \dots)$ and $(X'_n, X'_{n+1}, \dots)$ are a.s. identical, forming a single homogeneous Markov chain with transition probabilities p starting at z, and*

(c) *the two pre-T processes and the single post-T process are mutually independent. Under this probability $\tilde{P}$ the two marginal processes (X_n) and (X'_n) are Markov (μ, p) and Markov (μ', p) respectively; these chains agree $\tilde{P}$* a.s. *after the random time T when they first meet,*

$$\tilde{P}(T>n)=\|\alpha_n\|,$$

and the maximal coupling thus provided is the maximal coupling of Griffeath [4].

The existence and uniqueness of the $\tilde{P}$ described in the theorem is quite obvious, and it is also clear that $\tilde{P}$ makes the two marginal processes (X_n) and (X'_n) agree as soon as they have met at time T. Less obvious, however, is the fact that $\tilde{P}$ makes these marginal processes Markov (μ, p) and Markov (μ', p), though this is implied by the identification with Griffeath's maximal coupling, which is easily made. In view of the difficulty of Griffeath's construction a new proof is provided below of the fact that $\tilde{P}$ induces the right marginal processes, and it is hoped that this argument provides some insight into the way the coupling works. The argument depends on the fact that $\tilde{P}$ makes the random time T a *randomized stopping time* of each of the two marginal chains. This seems to be a feature of all manageable couplings of Markov chains, and the construction of $\tilde{P}$ is presented as an instance of a general method for constructing couplings of this type.

To keep the notation simple it is assumed throughout that we are dealing with homogeneous Markov chains with countable state space, but only the Markov property is really essential. Everything can easily be translated to apply to Markov chains with general measurable state space, and the results even extend to apply to inhomogeneous chains. The results in this case can be read off from those in the homogeneous case by using the device of the space-time chain.

2. Randomized Stopping Times and Couplings

Let (X_n) be a Markov (μ, p) chain defined on $(\Omega, \mathscr{F}, P)$. A random variable T defined on $(\Omega, \mathscr{F})$ with values in the extended time set $N \cup \{\infty\}$ is said to be a *randomized stopping time* (*r.s.t.*) *of* (X_n) if for each $n \in N$ the event $(T > n)$ is conditionally independent of the future $(X_n, X_{n+1}, \ldots)$ given the past $(X_0, \ldots, X_n)$. For motivation of this definition and proofs of the facts about r.s.t.'s which are now stated, see Pitman and Speed [12].

The most general possible joint distribution for (X_n) and a r.s.t. T of (X_n) is obtained by specifying a decreasing sequence of functions (f_n), with $f_n: S^n \to [0,1]$ such that for each sequence $x_0, x_1, \ldots$ of points in S, $1 \geqq f_0(x_0) \geqq f_1(x_0, x_1) \geqq \cdots \geqq 0$, and requiring that

$$P(T > n | X_0, X_1, \ldots) = f_n(X_0, \ldots, X_n), \qquad n \in N. \tag{5}$$

One way to achieve this is to let $(\Omega, \mathscr{F}, P)$ support both (X_n) and a random variable U independent of (X_n) with uniform distribution on $[0,1]$, and set

$$T = \inf\{n: f_n(X_0, \ldots X_n) \leqq U\}.$$

This construction, due to Wald and Wolfowitz [15], represents the r.s.t. T of (X_n) as a uniform mixture of stopping times of (X_n).

Given a r.s.t. T of a Markov (μ, p) chain (X_n), let $\mathscr{F}_n^T$ denote the σ-field generated by $X_0, \ldots, X_n$ and the events $(T=0), \ldots, (T=n)$. Then (X_n) is Markov with respect to $(\mathscr{F}_n^T)$, i.e., for $n, m \in N$, $y, z \in S$, $A \in \mathscr{F}_n^T$,

$$P(X_{n+m} = z | A, X_n = y) = p^m(y, z), \tag{6}$$

a formula which is particularly useful for $A = (T = n)$ and $A = (T > n)$. As a consequence of (6) a version of the strong Markov property for r.s.t.'s is easily formulated.

We shall make use of the following simple generalization of the inequality (3):

Lemma. *Let X and X' be S valued random variables defined on probability spaces $(\Omega, \mathscr{F}, P)$ and $(\Omega', \mathscr{F}', P')$, with distributions λ and λ' on S, and suppose there are events $F \in \mathscr{F}$ and $F' \in \mathscr{F}'$ such that $P(F) = P'(F')$ and the P distribution of X conditional on $\Omega \setminus F$ is identical to the P' distribution of X' conditional on $\Omega' \setminus F'$. Then*

$$\|\lambda - \lambda'\| \leqq P(F) = P(F'). \tag{7}$$

Proof. For $A \subset S$

$$\lambda(A) = P(F, X \in A) + P(\Omega \setminus F, X \in A),$$

$$\lambda'(A) = P'(F', X' \in A) + P'(\Omega' \setminus F', X' \in A).$$

But the assumptions on F and F' make $P(\Omega \setminus F, X \in A) = P'(\Omega' \setminus F', X' \in A)$, whence $|\lambda(A) - \lambda'(A)| = |P(F, X \in A) - P'(F', X' \in A)| \leqq P(F) = P'(F')$.

Proposition. *Let μ and μ' be two initial distributions on S, and set $\mu_n = \mu\, p^n$, $\mu'_n = \mu'\, p^n$. Let T be a r.s.t. of a Markov (μ, p) chain (X_n) defined on $(\Omega, \mathscr{F}, P)$, and let T' be a r.s.t. of a Markov (μ', p) chain (X'_n) defined on $(\Omega', \mathscr{F}', P')$. If the P distribution of (T, X_T) is identical to the P' distribution of $(T', X'_{T'})$, i.e., if*

$$P(T = n, X_n = y) = P'(T' = n, X'_n = y), \qquad n \in N,\ y \in S, \tag{8}$$

then

$$\|\mu_n - \mu'_n\| \leqq P(T > n) = P'(T' > n), \qquad n \in N. \tag{9}$$

Proof. It is easy to see that (8) and (6) imply

$$P(T \leqq n, X_n = y) = P'(T' \leqq n, X'_n = y), \qquad n \in N,\ y \in S,$$

and (9) now follows by the lemma.

The present inequality (9) can clearly be used in the same way as the coupling inequality (4) to derive ergodicity properties such as (1) and (2). Indeed, the connection between these two bounds is extremely close. If the random time T in (4) is a r.s.t. of each of the chains (X_n) and (X'_n) being considered there, then (4) is a special case of (9) with $(\Omega', \mathscr{F}', P') = (\Omega, \mathscr{F}, P)$, $T' = T$ and $X'_{T'} = X'_T = X_T$ (so that (8) is automatic). This is the situation for all the examples of couplings considered by Griffeath in [4], and in particular, for his maximal coupling described below. In general, (4) is not a sequence of (9), since it is possible to construct some rather bizarre couplings where T is the first time the two chains meet but where T is not a r.s.t. of either chain. However, these couplings seem to be of little interest since the maximal coupling shows that their use is never necessary.

On the other hand, given the set-up for (9) with the matching distributions (8), one can always splice things together to obtain a coupling which does just as well. Indeed, let $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{P})$ be a probability space on which there are defined a Markov (μ, p) chain $(\hat{X}_n)$, a Markov (μ', p) chain $(\hat{X}'_n)$, a random time $\hat{T}$ and an S valued random variable $\hat{Y}$ meeting the following mutually consistent requirements:

(i) *the $\hat{P}$ distribution of $[(\hat{X}_n), \hat{T}, \hat{Y}]$ equals the P distribution of $[(X_n), T, X_T]$;*

(ii) *the $\hat{P}$ distribution of $[(\hat{X}'_n, \hat{T}, \hat{Y}]$ equals the P' distribution of $[(X'_n, T', X'_{T'}]$;* (10)

(iii) *under $\hat{P}$ the Markov chains $(\hat{X}_n)$ and $(\hat{X}'_n)$ are conditionally independent given $\hat{T}$ and $\hat{Y}$.*

Notice that (i) and (ii) above make

$$\hat{Y} = \hat{X}_{\hat{T}} = \hat{X}'_{\hat{T}} \quad \hat{P} \text{ a.s.} \tag{11}$$

Now cross over from $(\hat{X}'_n)$ to $(\hat{X}_n)$ at time $\hat{T}$ to form $(\hat{X}''_n)$:

$$\begin{aligned} \hat{X}''_n &= \hat{X}'_n \quad \text{on } (\hat{T} > n) \\ &= \hat{X}_n \quad \text{on } (\hat{T} \leqq n). \end{aligned} \tag{12}$$

Then the conditional independence in (10(iii)) and the strong Markov property of the chains $(\hat{X}_n)$ and $(\hat{X}'_n)$ at the time $\hat{T}$ (which is a r.s.t. of each of these chains) imply that $(\hat{X}''_n)$ is Markov (μ', P) just like $(\hat{X}'_n)$. Thus $\hat{T}$ provides a coupling of the Markov (μ, p) chain $(\hat{X}_n)$ and the Markov (μ', p) chain $(\hat{X}''_n)$ which yields the same inequality in (4) as in (9).

We now make a construction which shows that the inequality (7) is sharp. When this construction is transformed into a coupling in the manner just described, it is Griffeath's maximal coupling which results.

Fix the initial distributions μ and μ', set $\alpha_0 = \mu - \mu'$,

$$\alpha_n = \mu_n - \mu'_n = \alpha_0 \, p^n,$$

and note that for $n \geqq 1$,

$$\alpha_n^+ = (\alpha_{n-1} \, p)^+ \leqq \alpha_{n-1}^+ \, p, \qquad \alpha_n^- = (\alpha_{n-1} \, p)^- \leqq \alpha_{n-1}^- \, p. \tag{13}$$

We first observe that to obtain equality in (9) it suffices to make the r.s.t. T of (X_n) and the r.s.t. T' of (X'_n) such that

$$P(T > n, X_n = y) = \alpha_n^+(y), \qquad P'(T' > n, X'_n = y) = \alpha_n^-(y), \qquad y \in S. \tag{14}$$

Not only does (14) imply equality in (9), but (14) also makes (8) automatic, since using (6) and (14) are finds that for $n \geqq 1$, $y \in S$

$$\mathrm{P}(\mathrm{T} = \mathrm{n}, \mathrm{X}_n = \mathrm{y}) = (\alpha_{n-1}^+ \, p - \alpha_n^+)(y) = (\alpha_{n-1}^- \, p - \alpha_n^-)(y) = P'(T' = n, X'_n = y), \tag{15}$$

the same being true for $n = 0$ if one replaces $\alpha_{n-1}^+ \, p$ by μ, $\alpha_{n-1}^- \, p$ by μ'. It now only remains to engineer (14). The simplest thing to try is to make the conditional probability of $(T > n)$ given $X_0, \ldots, X_n$ and $(T \geqq n)$ equal to $r_n(X_n)$ for some function $r_n \colon S \to [0, 1]$, i.e., to require (5) with

$$f_n(x_0, \ldots, x_n) = r_0(x_0) \, r_1(x_1) \ldots r_n(x_n), \tag{16}$$

doing a similar thing with the primed quantities. But a simple induction now shows to achieve (14) with this prescription one just has to take

$$r_0 = \alpha_0^+/\mu, \; r'_0 = \alpha_0^-/\mu'; \; r_n = \alpha_n^+/\alpha_{n-1}^+ \, p, \; r'_n = \alpha_n^-/\alpha_{n-1}^- \, p, \qquad n \geqq 1, \tag{17}$$

where for two measures β and γ on S with $\beta \leqq \gamma$, β/γ denotes the density of β with respect to γ, defined as a function on S by

$$(\beta/\gamma)(x) = \beta(x)/\gamma(x) \quad \text{if } \gamma(x) > 0, 0 \text{ otherwise.}$$

A feature of the construction just completed is that the r.s.t. T of (X_n) is such that conditional on either $(T = n)$ or $(T \geqq n)$, for each $n \geqq 1$, the pre-n process $(X_0, \ldots, X_n)$ is an inhomogeneous Markov chain. This is an easy consequence of the multiplicative structure (16) in the conditional distribution of T given (X_n), and, indeed, one finds that for either conditioning event, regardless of the value of $n \geqq 1$, the reverse transition probabilities of the inhomogeneous Markov chain are simply given as follows: the transition probability from y at time m to x at time $m-1$ is

$$\alpha_{m-1}^+(x) \, p(x, y)/\alpha_{m-1}^+ \, p(y). \tag{17}$$

Replacing + by − in (17) gives the corresponding conditional transition probability for the primed process.

Putting things together now we have:

Proof of the Theorem. Consider the process $((\hat{X}_n, \hat{X}''_n), n \in N)$ defined on $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{P})$ which is obtained in (10) and (12) from the set-up just described for equality in (9). It follows from (14) that the two pre-$\hat{T}$ processes cannot meet before $\hat{T}$ except on a set of $\hat{P}$-measure zero, so in view of (11) we must have

$$\hat{T} = \inf\{n : \hat{X}_n = \hat{X}''_n\}, \ \hat{P} \text{ a.s.}$$

Thus if we now let P^* denote the law on the double sequence space $(\tilde{\Omega}, \tilde{\mathscr{F}})$ of the process $((\hat{X}_n, \hat{X}''_n), n \in N)$, the preceding results transfer by change of variables to identify P^* with the $\tilde{P}$ described in the theorem, and all the asserted properties of $\tilde{P}$ follow immediately.

Notes. (i) Minor modifications extend the theorem to apply also in the case when either μ and μ' are not mutually singular or $\|\alpha_n\|$ does not converge to zero as $n \to \infty$.

(ii) Further distributional properties of $\tilde{P}$ can be read off by change of variables in (15), (16), and (17).

Acknowledgements. I would like to thank Terry Speed for his supervision of my research at the University of Sheffield in 1973, when the method of matching distributions at r.s.t.'s was developed. I also want to thank David Griffeath for some stimulating correspondence which led to the link-up with the coupling method, and thanks are due as well to David Freedman, Martin Jacobsen, and Michael Klass for their helpful suggestions.

References

1. 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 (1933)
2. Freedman, D.: Markov chains. San Francisco: Holden-Day 1971
3. Freedman, D.: The Poisson approximation for dependent events. Ann. Probability **2**, 256–269 (1974)
4. Griffeath, D.: A maximal coupling for Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 95–106 (1975)
5. Griffeath, D.: Coupling methods for Markov processes. Thesis, Cornell University (1975)
6. Griffeath, D.: Uniform coupling of nonhomogeneous Markov chains. [To appear in J. Appl. Probability (1975)]
7. Griffeath, D.: Partial coupling and loss of memory for Markov chains. [To appear in Ann. Probability (1976)]
8. Hodges, J.L., LeCam, L.: The Poisson approximation to the Poisson binomial distribution. Ann. Math. Statist. **31**, 737–740 (1960)
9. Jacobsen, M., Pitman, J.W.: Birth, death, and conditioning of Markov chains. [Submitted to Ann. Probability]
10. Orey, S.: Limit theorems for Markov chains. London: Van Nostrand, 1971
11. Pitman, J.W.: Uniform rates of convergence for Markov chain transition probabilities. Z. Wahrscheinlichkeitstheorie verw. Gebiete **29**, 193–227 (1974)
12. Pitman, J.W., Speed, T.P.: A note on random times. Stochastic Processes and Their Applications **1**, 369–374 (1973)

13. Serfling, R.J.: A general Poisson approximation theorem. Ann. Probability **3**, 726–731 (1975)
14. Vasershtein, L.N.: Markov processes on countable product spaces describing large systems of automata. Problemy Peredavci Informacii **3**, 64–72 (1969)
15. Wald, A. and Wolfowitz, J.: Two methods of randomization in statistics and the theory of games. Ann. of Math. **53**, 581–586 (1951)
16. Williams, D.: Path decomposition and continuity of local time for one-dimensional diffusions. Proc. London Math. Soc. 3rd Ser. **28**, 738–768 (1975)

Received February 2, 1976