Faithful couplings of Markov chains: now equals forever

by

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 9509, June 26, (1995)

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Faithful couplings of Markov chains: now equals forever

by

Jeffrey S. Rosenthal*

Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Phone: (416) 978-4594; Internet: jeff@utstat.toronto.edu

(June, 1995.)

Acknowledgements. I am very grateful to Gareth Roberts and Richard Tweedie for many insightful and inspiring discussions about these and related topics.

1. Introduction.

This short note considers the usual coupling approach to bounding convergence of Markov chains. It addresses the question of whether it suffices to have two chains become equal at a *single* time, or whether it is necessary to have them then *remain* equal for all future times.

Let $P(x, \cdot)$ be the transition probabilities for a Markov chain on a state space $\mathcal{X}$. Let μ and ν be two initial distributions for the chain. We are interested in bounding the total variation distance $\|\mu P^k - \nu P^k\| = \sup_{A \subseteq \mathcal{X}} |\mu P^k(A) - \nu P^k(A)|$ after k steps between the chain started in these two initial distributions.

(Often ν will be taken to be a stationary distribution for the chain, so that $\nu P^k = \nu$ for all $k \geq 0$. The problem then becomes one of convergence to stationarity for the Markov chain when started in the distribution μ. This is an important question for Markov chain Monte Carlo algorithms; see Gelfand and Smith, 1990; Smith and Roberts, 1993; Meyn and Tweedie, 1994; Rosenthal, 1995.)

The standard coupling approach to this problem (see Pitman, 1976; Lindvall, 1992; Thorisson, 1992) is as follows. We jointly define random variables X_k and Y_k, for $k = 0, 1, 2, \ldots$, such that $\mathcal{L}(X_0) = \mu$ and $\mathcal{L}(Y_0) = \nu$, and such that $\{X_k\}$ and $\{Y_k\}$ each *marginally* follow the transition probabilities for the Markov chain (i.e. $\mathbf{Pr}(X_{k+1} \in A \,|\, X_0 =$

* Partially supported by NSERC of Canada.

$x_0, \ldots, X_k = x_k) = P(x_k, A)$ and $\mathbf{Pr}(Y_{k+1} \in A \,|\, Y_0 = y_0, \ldots, Y_k = y_k) = P(y_k, A)$, for any $A \subseteq \mathcal{X}$ and any choices of $x_i, y_j \in \mathcal{X}$). It then follows that $\mathcal{L}(X_k) = \mu P^k$ and $\mathcal{L}(Y_k) = \nu P^k$, so that if T is a random time with

$$X_k = Y_k \qquad \text{for all} \quad k \geq T, \tag{$*$}$$

then the *coupling inequality* gives that

$$\|\mu P^k - \nu P^k\| = \|\mathcal{L}(X_k) - \mathcal{L}(Y_k)\| \leq \mathbf{Pr}(X_k \neq Y_k) \leq \mathbf{Pr}(T > k)\,.$$

This technique has been successfully applied to give useful bounds on distance to stationarity for a large number of examples; see for example Aldous (1983). We note that it is not required that the processes $\{X_k\}$ and $\{Y_k\}$ proceed independently; indeed, it is desired to define them jointly so that their probability of becoming equal to each other is as large as possible.

The purpose of this note is to examine when condition $(*)$ can be replaced by the simpler condition that $X_T = Y_T$, i.e. that the two chains become equal at some *one* time without necessarily remaining equal for all future times. Typically, given such a construction, one defines a new process $\{Z_k\}$ by

$$Z_k = \begin{cases} Y_k, & k \leq T \\ X_k, & k > T\,. \end{cases} \tag{$**$}$$

If one can show that $\{Z_k\}$ also marginally follows the transition probabilities for the Markov chain, then one can proceed as above. The purpose of this note is to provide sufficient conditions ("faithful couplings") under which this last step is automatic. It is to be hoped that this will simplify coupling arguments in future examples.

Remark. This work arose out of discussions with Richard Tweedie concerning bounding quantities like $\|P^n(\alpha, \cdot) - P^{n-1}(\alpha, \cdot)\|$, which arise in Meyn and Tweedie (1994). One possibility was to use coupling by choosing $X_0 = \alpha$, letting $\{X_k\}$ proceed according to the Markov chain, and then letting T be the smallest time at which $\{X_k\}$ stays still for one step, i.e. we have $X_{T+1} = X_T$. One could then set

$$Y_k = \begin{cases} X_k, & k \leq T \\ X_{k-1}, & k > T\,. \end{cases}$$

If it were true that $\{Y_k\}$ marginally followed the transition probabilities, then we would have

$$\|P^n(\alpha,\cdot) - P^{n-1}(\alpha,\cdot)\| = \|\mathcal{L}(Y_n) - \mathcal{L}(X_{n-1})\| \leq \mathbf{Pr}(Y_n \neq X_{n-1}) \leq \mathbf{Pr}(T > n).$$

However, this will <u>not</u> be the case in general. (For example, the first time k for which $Y_{k+1} = Y_k$ will be stochastically too large.) Such considerations motivated the current work.

2. Faithful couplings.

Given a Markov chain $P(x,\cdot)$ on a state space $\mathcal{X}$, we define a *faithful coupling* to be a collection of random variables X_k and Y_k for $k \geq 0$, defined jointly on the same probability space, such that

$$(i) \quad \mathbf{Pr}(X_{k+1} \in A \,|\, U_k = u, X_0 = x_0, \ldots, X_k = x_k, Y_0 = y_0, \ldots, Y_u = y_u) = P(x_k, A)$$

and

$$(ii) \quad \mathbf{Pr}(Y_{k+1} \in A \,|\, U_k = u, X_0 = x_0, \ldots, X_u = x_u, Y_0 = y_0, \ldots, Y_k = y_k) = P(y_k, A)$$

for all $A \subseteq \mathcal{X}$ and $x_i, y_j \in \mathcal{X}$, where

$$U_k = \min(k, \inf\{j \geq 0\,;\, X_j = Y_j\}).$$

Intuitively, a faithful coupling is one in which the influence of each chain upon the other is not too great.

To check faithfulness, it suffices to check it with U_k is replaced by k, in which case it becomes equivalent to the following two conditions:

(a) the pairs process $\{(X_k, Y_k)\}_{k=1}^{\infty}$ is a Markov chain on $\mathcal{X} \times \mathcal{X}$;

(b) for any $k \geq 0$ and $x_k, y_k \in \mathcal{X}$, and for any measurable subset $A \subseteq \mathcal{X}$,

$$\mathbf{Pr}(X_{k+1} \in A \,|\, X_k = x_k, Y_k = y_k) = P(x_k, A)$$

and

$$\mathbf{Pr}(Y_{k+1} \in A \,|\, X_k = x_k, Y_k = y_k) = P(y_k, A).$$

Here condition (a) merely says that the coupling is jointly Markovian; condition (b) says that the updating probabilities for one process are not affected by the previous value of the other process. Both of these conditions are satisfied (and easily verified) in many different couplings used in specific examples. This is the case, e.g., for couplings defined by minorization conditions (Meyn and Tweedie, 1993; Rosenthal, 1995).

In this section we prove that for faithful couplings, it suffices, in bounding convergence, that the two processes becoming equal at some *one* time. In the next section, we provide an example to show that in general condition (a) alone is not sufficient to allow the construction given by (**) above.

Theorem 1. *Given a Markov chain $P(x,\cdot)$ on a state space $\mathcal{X}$, let $\{X_k, Y_k\}_{k=0}^{\infty}$ be a faithful coupling as defined above. Set $\mu = \mathcal{L}(X_0)$ and $\nu = \mathcal{L}(Y_0)$, and let*

$$T = \inf\{k \geq 0\,;\, X_k = Y_k\}\,. \qquad (***)$$

Then

$$\|\mu P^k - \nu P^k\| \leq \mathbf{Pr}(T > k)\,.$$

Proof. As in (**), we define

$$Z_k = \begin{cases} Y_k, & k \leq T \\ X_k, & k > T\,. \end{cases}$$

If we can show that $\{Z_k\}$ marginally follows the transition probabilities $P(x,\cdot)$, then it will follow that $\mathcal{L}(Z_k) = \nu P^k$, so that from the coupling inequality, we would have

$$\|\mu P^k - \nu P^k\| = \|\mathcal{L}(X_k) - \mathcal{L}(Z_k)\| \leq \mathbf{Pr}(X_k \neq Z_k) = \mathbf{Pr}(T > k)\,,$$

giving the result.

To proceed, we note that

$$\mathbf{Pr}(Z_{k+1} \in A, Z_0 \in dz_0, \ldots, Z_k \in dz_k)$$

$$= \mathbf{Pr}(Z_{k+1} \in A, Z_0 \in dz_0, \ldots, Z_k \in dz_k, T \geq k+1)$$

$$+\sum_{t=0}^{k}\mathbf{Pr}(Z_{k+1}\in A, Z_0\in dz_0,\ldots,Z_k\in dz_k, T=t)$$

$$=\int_{\substack{x_0,\ldots,x_k\in\mathcal{X}\\ x_i\neq z_i,\ 0\le i\le k}}\mathbf{Pr}(Y_{k+1}\in A, X_0\in dx_0,\ldots,X_k\in dx_k, Y_0\in dz_0,\ldots,Y_k\in dz_k)$$

$$+\sum_{t=0}^{k}\int_{\substack{x_0,\ldots,x_{t-1}\in\mathcal{X}\\ x_i\neq z_i,\ 0\le i\le t-1}}\mathbf{Pr}(X_{k+1}\in A, X_0\in dx_0,\ldots,X_{t-1}\in dx_{t-1},$$

$$X_t\in dz_t,\ldots,X_k\in dz_k, Y_0\in dz_0,\ldots,Y_t\in dz_t)\,.$$

Using conditions (i) and (ii) above, this is equal to

$$\int_{\substack{x_0,\ldots,x_k\in\mathcal{X}\\ x_i\neq z_i,\ 0\le i\le k}}P(z_k,A)\ \mathbf{Pr}(X_0\in dx_0,\ldots,X_k\in dx_k, Y_0\in dz_0,\ldots,Y_k\in dz_k)$$

$$+\sum_{t=0}^{k}\int_{\substack{x_0,\ldots,x_{t-1}\in\mathcal{X}\\ x_i\neq z_i,\ 0\le i\le t-1}}P(z_k,A)\ \mathbf{Pr}(X_0\in dx_0,\ldots,X_{t-1}\in dx_{t-1},$$

$$X_t\in dz_t,\ldots,X_k\in dz_k, Y_0\in dz_0,\ldots,Y_t\in dz_t)$$

$$=P(z_k,A)\ \mathbf{Pr}(Z_0\in dz_0,\ldots,Z_k\in dz_k, T\ge k+1)$$

$$+\,P(z_k,A)\ \sum_{t=0}^{k}\mathbf{Pr}(Z_0\in dz_0,\ldots,Z_k\in dz_k, T=t)$$

$$=P(z_k,A)\mathbf{Pr}(Z_0\in dz_0,\ldots,Z_k\in dz_k)\,.$$

It follows that $\{Z_k\}$ marginally follows the transition probabilities $P(x,\cdot)$, as required. ∎

Corollary 2. *Let $\{X_k,Y_k\}$ be a faithful coupling, and let T' be any stopping time with $X_{T'}=Y_{T'}$. Then*

$$\|\mu P^k-\nu P^k\|\le\mathbf{Pr}(T'>k)\,.$$

Proof. We clearly have $T \leq T'$. Hence, $\|\mu P^k - \nu P^k\| \leq \mathbf{Pr}(T > k) \leq \mathbf{Pr}(T' > k)$. ∎

Remark. Analogous results to the above clearly hold for continuous-time Markov processes as well.

3. An example concerning necessity.

We present here an example to show that condition (a) alone is not sufficient for the above construction.

Proposition 3. *There exists a Markov chain $P(x, \cdot)$ on a state space $\mathcal{X}$, and a Markovian coupling $\{(X_k, Y_k)\}$ on $\mathcal{X} \times \mathcal{X}$ with each of $\{X_k\}$ and $\{Y_k\}$ marginally following the transition probabilities $P(x, \cdot)$, such that if T is defined by $(***)$, and the process $\{Z_k\}$ is then defined by $(**)$, then the process $\{Z_k\}$ will not marginally follow the transition probabilities $P(x, \cdot)$.*

Proof. Let $\mathcal{X} = \{0, 1\}$. Define a Markov chain on $\mathcal{X}$ by $P(0,0) = P(1,1) = P(0,1) = P(1,0) = \frac{1}{2}$. (That is, this is the Markov chain corresponding to i.i.d. choices at each time.)

Define a Markov chain on $\mathcal{X} \times \mathcal{X}$ as follows. At time 0, take $\mathbf{Pr}(Y_0 = 0) = \mathbf{Pr}(Y_0 = 1) = \mathbf{Pr}(X_0 = 0) = \mathbf{Pr}(X_0 = 1) = \frac{1}{2}$. For $k \geq 0$, conditional on $(X_k, Y_k) \in \mathcal{X} \times \mathcal{X}$, we let

$$\mathbf{Pr}(Y_{k+1} = 0 \,|\, X_k, Y_k) = \mathbf{Pr}(Y_{k+1} = 1 \,|\, X_k, Y_k) = \frac{1}{2},$$

and set

$$X_{k+1} \;=\; X_k \oplus Y_k \;\equiv\; X_k + Y_k \pmod 2\,.$$

In words, $\{Y_k\}$ proceeds according to the original Markov chain $P(x, \cdot)$, without regard to the values of $\{X_k\}$. On the other hand, $\{X_k\}$ changes values precisely when the corresponding value of $\{Y_k\}$ is 1.

It is easily verified that, for all $k \geq 0$, we will have $\{Y_k\}$ i.i.d. equal to 0 or 1 with probability $\frac{1}{2}$. Using this, it is easily verified that $\{X_k\}$ will be similarly i.i.d.

Hence, $\{(X_k, Y_k)\}$ is a Markovian coupling, with each coordinate marginally following the chain $P(x, \cdot)$. On the other hand, conditions (i) and (b) above are clearly violated.

Now, letting $T = \inf\{k \geq 0\,;\, X_k = Y_k\}$, and defining $\{Z_k\}$ as in $(**)$, we have that

$$\begin{aligned}\mathbf{Pr}(Z_1 = 1, Z_0 = 0) &= \mathbf{Pr}(Z_1 = 1, Z_0 = 0, T > 0) + \mathbf{Pr}(Z_1 = 1, Z_0 = 0, T = 0) \\ &= \mathbf{Pr}(Y_1 = 1, Y_0 = 0, X_0 = 1) + \mathbf{Pr}(X_1 = 1, Y_0 = 0, X_0 = 0) \\ &= \frac{1}{8} + 0 \\ &= 1/8\,.\end{aligned}$$

It follows that $\mathbf{Pr}(Z_1 = 1 \,|\, Z_0 = 0) = 1/4$, which is not equal to $P(0,1) = 1/2$. ∎

4. Faithful shift-coupling.

A result analogous to Theorem 1 can also be proved for the related method of shift-coupling.

Given processes $\{X_k\}$ and $\{Y_k\}$, each marginally following the transition probabilities $P(x, \cdot)$, random times T and T' are called *shift-coupling epochs* (Aldous and Thorisson, 1993; Thorisson, 1992, Section 10) if $X_{T+k} = Y_{T'+k}$ for all $k \geq 0$. The shift-coupling inequality (Thorisson, 1992, equation 10.2; Roberts and Rosenthal, 1994, Proposition 1) then gives that

$$\left\| \frac{1}{n} \sum_{k=1}^{n} \mathbf{Pr}(X_k \in \cdot) \;-\; \frac{1}{n} \sum_{k=1}^{n} \mathbf{Pr}(X'_k \in \cdot) \right\| \leq \frac{1}{n} \mathbf{E}\left(\min\left(\max(T, T'), n\right)\right) .$$

The quantity $\max(T, T')$ thus serves to bound the difference of the average distributions of the two chains.

For shift-coupling, since we will not in general have $T = T'$, the definition of faithful given above is not sufficient. Thus, we define a collection of random variables $\{X_k, Y_k\}$ to be a *faithful shift-coupling* if we have

$$(i') \qquad \mathbf{Pr}(X_{k+1} \in A \,|\, R_k = r, X_0 = x_0, \ldots, X_k = x_k, Y_0 = y_0, \ldots, Y_r = y_r) \;=\; P(x_k, A)$$

and

$$(ii') \qquad \mathbf{Pr}(Y_{k+1} \in A \,|\, S_k = s, X_0 = x_0, \ldots, X_s = x_s, Y_0 = y_0, \ldots, Y_k = y_k) \;=\; P(y_k, A)$$

for all $A \subseteq \mathcal{X}$ and $x_i, y_i \in \mathcal{X}$, where

$$R_k = \inf\{j \geq 0\,;\, \exists i \leq k, Y_j = X_i\}\,; \qquad S_k = \inf\{i \geq 0\,;\, \exists j \leq k, X_i = Y_j\}\,.$$

If $\{X_k, Y_k\}$ is a faithful shift-coupling, then the following theorem (cf. Roberts and Rosenthal, 1994, Corollary 3) shows that it suffices to have $X_T = Y_{T'}$ for some specific pair of times T and T'.

Theorem 4. *Let $\{X_k, Y_k\}$ be a faithful shift-coupling, and let*

$$\tau = \inf\{k \geq 0\,;\, \exists i,j \leq k \text{ with } X_i = Y_j\}\,.$$

Then

$$\left\| \frac{1}{n}\sum_{k=1}^{n} \mathbf{Pr}(X_k \in \cdot) \;-\; \frac{1}{n}\sum_{k=1}^{n} \mathbf{Pr}(X'_k \in \cdot) \right\| \leq \frac{1}{n}\mathbf{E}\left(\min\left(\tau, n\right)\right)\,.$$

Proof. Let $I, J \leq \tau$ be random times with $X_I = Y_J$. By minimality of τ, we must have $\max(I, J) = \tau$. Define $\{Z_k\}$ by

$$Z_k \;=\; \begin{cases} Y_k, & k \leq J \\ X_{k-J+I}, & k > J\,. \end{cases}$$

Using properties (i') and (ii') above, and summing over possible values of I and J and of intermediate states, it is checked as in Theorem 1 that $\{Z_k\}$ marginally follows the transition probabilities $P(x, \cdot)$. Hence $\mathcal{L}(Z_k) = \mathcal{L}(Y_k)$. Furthermore, the times $T = I$ and $T' = J$ are shift-coupling epochs for $\{X_k\}$ and $\{Z_k\}$. Since $\max(T, T') = \tau$, the result follows from the shift-coupling inequality applied to $\{X_k\}$ and $\{Z_k\}$. ∎

Corollary 5. *Let $\{X_k, Y_k\}$ be a faithful shift-coupling, and let T and T' be random times with $X_T = Y_{T'}$. Then*

$$\left\| \frac{1}{n}\sum_{k=1}^{n} \mathbf{Pr}(X_k \in \cdot) \;-\; \frac{1}{n}\sum_{k=1}^{n} \mathbf{Pr}(X'_k \in \cdot) \right\| \leq \frac{1}{n}\mathbf{E}\left(\min\left(\max(T, T'), n\right)\right)\,.$$

Proof. We clearly have $\tau \leq \max(T, T')$. The result follows. ∎

REFERENCES

D. Aldous (1983), Random walk on finite groups and rapidly mixing Markov chains. In Seminaire de Probabilités **XVII**, 243-297. Lecture notes in mathematics **986**.

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

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

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

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

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

G.O. Roberts and J.S. Rosenthal (1994), Shift-coupling and convergence rates of ergodic averages. Preprint.

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

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.

H. Thorisson (1992), Coupling methods in probability theory. Tech. Rep. RH-18-92, Science Institute, University of Iceland.