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; revised, August 1995.)

Acknowledgements. I am very grateful to Gareth Roberts and Richard Tweedie for many insightful 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 Polish state space $\mathcal{X}$. Let μ and ν be two initial distributions for the chain. This paper is related to the problem of 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 Doeblin, 1938; Griffeath, 1975; 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 $\{X_k\}$ is Markov (μ, P) (i.e. $\mathbf{Pr}(X_0 \in A) = \mu(A)$ and $\mathbf{Pr}(X_{k+1} \in A \,|\, X_0 = x_0, \ldots, X_k = x_k) = P(x_k, A)$ for any measurable

* Partially supported by NSERC of Canada.

$A \subseteq \mathcal{X}$ and any choices of $x_i \in \mathcal{X}$) and $\{Y_k\}$ is Markov (ν, P). 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

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

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; Lindvall, 1992). We emphasize 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.

When constructing couplings, instead of establishing condition $(*)$ directly, one often begins by establishing 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. Then, 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 $\{Z_k\}$ is again Markov (μ, P), then one can proceed as above, with $\{X_k\}$ replaced by $\{Z_k\}$. However, $\{Z_k\}$ will not be Markov (μ, P) in general; a simple counter-example is provided in Section 3.

The purpose of this note is to provide (Section 2) a fairly general condition ("faithfulness") under which the process $\{Z_k\}$ above will automatically be Markov (μ, P). We note that such issues are rather well studied, and we provide only a slight extension of previous ideas. Indeed, it has been argued by Pitman (1976, pp. 319-320) that $\{Z_k\}$ will be Markov (μ, P) provided that $\{X_n\}$ and $\{Y_n\}$ are conditionally independent, given T and X_T. However, this is a somewhat strong condition, and we shall provide an example (Section 3) of a common coupling which satisfies our faithfulness condition, but for which this conditional independence does not hold.

In Section 4, we shall show that our approach generalizes to a similar condition for the related method of shift-coupling.

Remarks.

1. This work arose out of discussions with Richard Tweedie concerning bounding quantities like $\|P^n(\alpha,\cdot) - P^{n-1}(\alpha,\cdot)\|$, where α is an atom, 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. for which $X_{T+1} = X_T$. One could then set

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

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

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

   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. If bounds on $\|\mu P^k - \nu P^k\|$ are the only item of interest, and an actual coupling is not required, then it may not be necessary to construct the process $\{Z_k\}$ at all. Indeed, the approach of *distributional* or *weak coupling* (Pitman, 1976, p. 319; Ney, 1981; Thorisson, 1983; Lindvall, 1992, §I.4) shows that if $\{(X_k, Y_k)\}$ is a process on $\mathcal{X} \times \mathcal{X}$, with $X_T = Y_T$, then if T is a randomized stopping time for each of $\{X_k\}$ and $\{Y_k\}$, or more generally if $(T, X_T, X_{T+1}, \ldots) \stackrel{\mathbf{d}}{=} (T, Y_T, Y_{T+1}, \ldots)$, then we have $\|\mathcal{L}(X_n) - \mathcal{L}(Y_n)\| \le \mathbf{Pr}(T > n)$, with no coupling construction necessary. (Note that these conditions on T will not always hold; cf. Pitman, 1976, p. 319, and also the example in the proof Proposition 3 below with, say, $T = 0$. But they will hold for couplings which are faithful.)

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

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

and

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

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

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

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 (because then we are conditioning on more information), 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 $A \subseteq \mathcal{X}$,

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

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 (see Section 3).

In this section we prove that for faithful couplings, the process $\{Z_k\}$ will automatically be Markov (μ, P). In the next section, we provide an example to show that in general condition (a) alone is not sufficient for this.

Theorem 1. *Given a Markov chain $P(x,\cdot)$ on a Polish 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*

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

Then the process $\{Z_k\}$, defined by $(**)$, is Markov (μ, P). (Hence, $\|\mu P^k - \nu P^k\| \leq \mathbf{Pr}(T > k)$.)

Proof. We first note (cf. Lindvall, 1992, p. 12) that, since $\mathcal{X}$ is assumed to be Polish, sets of the form $\{X_k = Y_k\}$ are measurable, and hence T is a well-defined random variable.

As for the process $\{Z_k\}$, we clearly have $\mathcal{L}(Z_0) = \mu$. To proceed, we note that

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

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

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

$$+ \sum_{t=0}^{k} \int\limits_{\substack{x_0, \ldots, x_{t-1} \in \mathcal{X} \\ x_i \neq z_i,\ 0 \leq i \leq t-1}} \mathbf{Pr}(X_{k+1} \in A, X_0 \in dx_0, \ldots, X_{t-1} \in dx_{t-1},$$
$$X_t \in dz_t, \ldots, X_k \in dz_k, Y_0 \in dz_0, \ldots, Y_t \in dz_t)\,.$$

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

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

$$+ \sum_{t=0}^{k} \int\limits_{\substack{x_0, \ldots, x_{t-1} \in \mathcal{X} \\ x_i \neq z_i,\ 0 \leq i \leq t-1}} P(z_k, A)\ \mathbf{Pr}(X_0 \in dx_0, \ldots, X_{t-1} \in dx_{t-1},$$
$$X_t \in dz_t, \ldots, X_k \in dz_k, Y_0 \in dz_0, \ldots, Y_t \in dz_t)$$

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

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

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

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

We thus obtain the following corollary, which also follows from the distributional coupling method (see the final remark of the Introduction).

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

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

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. Examples.

We first present 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 <u>not</u> 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 <u>not</u> equal to $P(0,1) = 1/2$. ∎

We now mention an example of a coupling which is faithful, but which does not satisfy the simpler condition that the processes $\{X_k\}$ and $\{Y_k\}$ are conditionally independent, given T and X_T. Our example involves minorization conditions (Nummelin, 1984; Meyn and Tweedie, 1993; Rosenthal, 1995).

Specifically, suppose that for our transition probabilities $P(x,\cdot)$, there is a subset C, some $\epsilon > 0$, and a probability measure $Q(\cdot)$ on $\mathcal{X}$, such that the *minorization condition*

$$P(x,\cdot) \geq \epsilon\, Q(\cdot)\,, \qquad x \in C$$

holds. Then we can define processes $\{X_k\}$ and $\{Y_k\}$ jointly as follows. Given X_n and Y_n,

(I) if $(X_n, Y_n) \in C \times C$, then flip an independent coin with probability of heads equal to ϵ. If the coin is heads, set $X_{n+1} = Y_{n+1} = q$ where $q \sim Q(\cdot)$ is chosen independently. If the coin is tails, choose $X_{n+1} \sim \frac{1}{1-\epsilon}\left(P(X_n,\cdot) - \epsilon\, Q(\cdot)\right)$ and $Y_{n+1} \sim \frac{1}{1-\epsilon}\left(P(Y_n,\cdot) - \epsilon\, Q(\cdot)\right)$, independently.

(II) if $(X_n, Y_n) \notin C \times C$, then choose $X_{n+1} \sim P(X_n,\cdot)$ and $Y_{n+1} \sim P(Y_n,\cdot)$, independently.

Then each of $\{X_n\}$ and $\{Y_n\}$ marginally follows the transition probabilities $P(x,\cdot)$. Furthermore the joint construction is easily seen to be faithful. Hence the results of the previous section apply. (In particular, we may let T' be the first time we choose option (I) above and the coin comes up heads.)

However, it is *not* the case that the two processes $\{X_n\}$ and $\{Y_n\}$ are conditionally independent given T and X_T. Indeed, given T and X_T, then for $n < T-1$ and $X_n \in C$, the conditional distribution of X_{n+1} may depend greatly on the event $\{Y_n \in C\}$. Thus, this is an example where the faithfulness condition holds, even though the simpler condition of conditional independence does not hold.

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 $\{Y_k\}$ on a state space $\mathcal{X}$, 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 on a Polish state space $\mathcal{X}$, 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**.

D.J. Aldous and H. Thorisson (1993), Shift-coupling. Stoch. Proc. Appl. **44**, 1-14.

W. Doeblin (1938), Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états. Rev. Math. Union Interbalkanique **2**, 77-105.

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

D. Griffeath (1975), A maximal coupling for Markov chains. Z. Wahrsch. verw. Gebiete **31**, 95-106.

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.

P. Ney (1981), A refinement of the coupling method in renewal theory. Stoch. Proc. Appl. **11**, 11-26.

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

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 (1983), The coupling of regenerative processes. Adv. Appl. Prob. **15**, 531-561.

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