# Shift-coupling and convergence rates of ergodic averages

by

Gareth O. ROBERTS
Statistical Laboratory
University of Cambridge, Cambridge CB2 1SB, U.K.
Internet: G.O.Roberts@statslab.cam.ac.uk

and

Jeffrey S. ROSENTHAL
Department of Statistics
University of Toronto, Toronto, Ontario, Canada M5S 3G3
Internet: jeff@utstat.toronto.edu

(To appear in *Communications in Statistics – Stochastic Models.*)

**Abstract.** We study convergence of Markov chains $\{X_k\}$ to their stationary distributions $\pi(\cdot)$. Much recent work has used coupling to get quantitative bounds on the total variation distance between the law $\mathcal{L}(X_n)$ and $\pi(\cdot)$. In this paper, we use shift-coupling to get quantitative bounds on the total variation distance between the ergodic average law $\frac{1}{n}\sum_{k=1}^{n}\mathcal{L}(X_k)$ and $\pi(\cdot)$. This avoids certain problems, related to periodicity and near-periodicity of the Markov chain, which have plagued previous work.

**Keywords.** shift-coupling, computable bounds, Markov chain Monte Carlo, drift condition, minorization condition, small set.

## 1. INTRODUCTION.

Recent work of Meyn and Tweedie (1994), Rosenthal (1995), and Baxendale (1994) has used minorization conditions and drift conditions to obtain quantitative bounds on convergence to stationarity of certain Markov chains. Such bounds are particularly important in the applied area of Markov chain Monte Carlo (see e.g. Gelfand and Smith, 1990; Smith and Roberts, 1993), as a method of determining how long to run a Markov chain until it can be regarded as approximately a sample from the desired stationary distribution.

Much of this theoretical analysis has used the method of *coupling* (see e.g. Lindvall, 1992), in which two different versions of the Markov chain (one started in the stationary distribution) are defined jointly in such a way that they are likely to be equal after a large number of iterations. The coupling inequality then immediately provides bounds on the distance to stationarity. One of the difficulties in applying such results is that the two Markov chains must become equal *at the same time*. This can be difficult to accomplish, and in particular it is plagued by difficulties related to periodicity (or near-periodicity) of the underlying Markov chain.

In this paper, we avoid such difficulties by examining convergence of ergodic averages $\frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot)$ of the Markov chain distributions (see e.g. Tierney, 1994, Section 3.3), rather than of the individual distributions $P(X_n \in \cdot)$ themselves. This allows us to replace the method of coupling with the related method of *shift-coupling* (see Aldous and Thorisson, 1993; Thorisson, 1992, Section 10; Thorisson, 1993; Thorisson, 1994), in which two Markov chains are allowed to become equal at *different* times. From the Monte Carlo perspective, this corresponds to sampling the chain at a time chosen uniformly in $\{1, 2, \ldots, n\}$ rather than at time $n$. This will sometimes be a less efficient sampling method, however it avoids all problems related to near-periodicity of the chain. (For other approaches to sampling from complicated distributions, see Asmussen, Glynn, and Thorisson, 1992; Propp and Wilson, 1995.)

Background and necessary results on shift-coupling are introduced in Section 2. Bounds using minorization and drift conditions, including a generalization involving multiple minorization conditions and also connections to potential theory, are presented in Section 3. Bounds for various examples are presented in Section 4.

## 2. BOUNDS INVOLVING SHIFT-COUPLING.

Let $P(\cdot,\cdot)$ be the transition probabilities for a Markov chain on a (possibly continuous) state space $\mathcal{X}$. Let $\{X_k\}_{k=0}^\infty$ and $\{X'_k\}_{k=0}^\infty$ be two processes defined jointly on $\mathcal{X}$, each marginally following the transition probabilities $P(\cdot,\cdot)$. Let $T$ and $T'$ be non-negative-extended-integer-valued random variables, such that if $T, T' < \infty$, then for each non-negative integer $n$,

$$X_{T+n} \;=\; X'_{T'+n}\,.$$

Then, following Aldous and Thorisson (1993), we call $T$ and $T'$ *shift-coupling epochs.*

Ordinary coupling requires that $T = T'$, in which case the total variation distance between the laws of $X_k$ and $X'_k$ satisfies

$$\|\mathcal{L}(X_k) \,-\, \mathcal{L}(X'_k)\| \;\equiv\; \sup_{A\subseteq\mathcal{X}} |P(X_k\in A) \,-\, P(X'_k\in A)| \;\le\; P(T>k)\,.$$

This is the well-known *coupling inequality.* Shift-coupling cannot possibly bound $\|\mathcal{L}(X_k) - \mathcal{L}(X'_k)\|$, since for example it cannot rule out periodic behaviour. However, it can be used to bound convergence of ergodic averages. The following bound is stated in Thorisson (1992, equation 10.2); for a proof see Thorisson (1993) or Thorisson (1994).

**Proposition 1.** *Let $\{X_k\}$, $\{X'_k\}$, $T$, and $T'$ be as above. Then the total variation distance between ergodic averages of $\{X_k\}$ and $\{X'_k\}$ satisfies*

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

**Remarks.**

1. This proposition would usually be used with $\mathcal{L}(X'_0) = \pi(\cdot)$, where $\pi$ is a stationary distribution for the chain. Then $\frac{1}{n}\sum_{k=1}^n P(X'_k\in A) \;=\; \pi(A)$. The proposition then provides a bound on the total variation distance between the stationary distribution $\pi(\cdot)$ and the average distribution $\frac{1}{n}\sum_{k=1}^n P(X_k\in\cdot)$ of the Markov chain of interest. Thus, if we can construct shift-coupling epochs, and can bound quantities of the form $P\left(\max(T,T')\ge \ell\right)$, then we can bound the distance to stationarity of ergodic averages.
2. The ergodic average of distributions $\frac{1}{n}\sum_{k=1}^n P(X_k\in\cdot)$ considered here has an interpretation related to estimating $\pi(A)$ by considering what fraction of the Markov chain states $X_1,\ldots,X_n$ lie in the set $A$. Such estimation is a common technique in Markov chain Monte Carlo. The above proposition then bounds (uniformly in the set $A$) the bias of this estimator, i.e. the absolute-value difference of the expected value of this estimator and the true quantity $\pi(A)$.

3. The ergodic average of distributions considered here, $\frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot)$, is of course different from the distribution of ergodic averages, $P\left(\frac{1}{n}\left(\sum_{k=1}^{n} X_k\right) \in \cdot\right)$. For the latter, few quantitative bounds are known, but various central limit theorems are available; see Geyer (1992) for a review. It is possible that shift-coupling could also be used in this context; for example, if $X_k, X'_k \in \mathbf{R}$ with $|X_k|, |X'_k| \leq C$ for all $k$, then

$$\left| \frac{1}{n}\left(\sum_{k=1}^{n} X_k\right) \; - \; \frac{1}{n}\left(\sum_{k=1}^{n} X'_k\right) \right| \; \leq \; \frac{2C}{n}\max(T, T')\,.$$

One possible method of constructing shift-coupling epochs is to allow $\{X_k\}$ and $\{X'_k\}$ to each marginally follow the Markov chain transitions $P(\cdot,\cdot)$, and to find times $T$ and $T'$ with $X_T = X'_{T'}$. We can then define $\{X''_k\}$ by $X''_k = X_k$ for $k \leq T$, and $X''_k = X'_{k-T+T'}$ for $k \geq T$. Then $T$ and $T'$ are shift-coupling epochs for $\{X''_k\}$ and $\{X'_k\}$, and $\mathcal{L}(X''_0) = \mathcal{L}(X_0)$. However, such a process $\{X''_k\}$ will not necessarily marginally follow the Markov chain transitions. We note that it suffices to have

$$\mathcal{L}\left(X'_{n+1}\,|\,T, T'_n, X_0, \ldots, X_T, X'_0, \ldots, X'_n\right) \; = \; P(X'_n, \cdot)\,, \quad n = 0, 1, 2, \ldots \qquad (*)$$

(where $T'_n = T'$ if $T' \leq n$, otherwise $T'_n = \infty$); this implies that $\mathcal{L}(X''_0, X''_1, \ldots) = \mathcal{L}(X_0, X_1, \ldots)$. [In particular, $(*)$ is satisfied if the chains are independent and the event $\{T = n, T' = n'\}$ is determined by $X_1, \ldots, X_n, X'_1, \ldots, X'_{n'}$.] In that case, we can apply our previous results to the chains $\{X''_k\}$ and $\{X'_k\}$, to obtain

**Corollary 2.** *Let $\{X_k\}$ and $\{X'_k\}$ each marginally follow the Markov chain transitions $P(\cdot,\cdot)$. Let $T$ and $T'$ be random times with $X_T = X'_{T'}$, and assume that $(*)$ is satisfied. Then*

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

One way to satisfy $(*)$ is to let $T' = 0$ always, let $T = \inf\{n \geq 0\,;\, X_n = X'_0\}$, and let $\{X'_k\}$ be conditionally independent of $T, X_0, \ldots, X_T$ given $X'_0$. Then $\max(T, T') = T$. Furthermore, on a *finite* state space, if $X'_0$ is distributed according to a stationary distribution $\pi(\cdot)$, then the expectation of $T$ does not depend on the starting point $X_0$ (Aldous, 1993, Chapter 2, Corollary 13). We thus obtain

**Corollary 3.** *Let $P(\cdot,\cdot)$ be the transition probabilities for a Markov chain on a finite state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$. Then uniformly over all starting distributions $\mathcal{L}(X_0)$, we have*

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) \; - \; \pi(\cdot) \right\| \;\leq\; \frac{1}{n}\, \mathbf{E}T \,,$$

*where $\mathbf{E}T = \sum_{j\in\mathcal{X}} \pi(j) \mathbf{E}_i(T_j)$ is the expected time, starting from the point $i \in \mathcal{X}$, to hit a point $j \in \mathcal{X}$ chosen according to $\pi(\cdot)$; furthermore, $\mathbf{E}T$ does not depend on $i$.*

Finally, we observe that most of this section applies equally well to continuous-time processes, with virtually identical proofs. We leave the details to the reader.

## 3. MINORIZATION AND DRIFT CONDITIONS.

Suppose our Markov chain, on a state space $\mathcal{X}$, has transition probabilities $P(\cdot,\cdot)$ which satisfy the following two inequalities, for some measurable function $V : \mathcal{X} \to [1,\infty)$, some probability measure $Q(\cdot)$ on $\mathcal{X}$, and some $\lambda < 1$, $\Lambda < \infty$, $\epsilon > 0$, and $d \geq 0$. Write $C = \{x \in \mathcal{X} \mid V(x) \leq d\}$.

(i) (*drift condition*) $E\left(V(X_1) \mid X_0 = x\right) \leq \lambda V(x) + \Lambda \mathbf{1}_C(x)$ for all $x \in \mathcal{X}$;

(ii) (*minorization condition*) $P(x,\cdot) \geq \epsilon\, Q(\cdot)$, for all $x \in C$.

These and related inequalities were used in Meyn and Tweedie (1994) and in Rosenthal (1995) to construct couplings and thus obtain bounds on the distance from stationarity of the individual distributions $\mathcal{L}(X_n)$. We use them here to construct shift-couplings and obtain simple bounds on the distance from stationarity of ergodic averages $\frac{1}{n}\sum_{k=1}^{n} \mathcal{L}(X_k)$ of the distributions.

Indeed, we shall define processes $\{X_k\}$ and $\{X'_k\}$ as follows. We choose $X_0$ according to an initial distribution $\nu(\cdot)$, and choose $X'_0$ independently according to a stationary distribution $\pi(\cdot)$. We let $\{X_k\}$ proceed as follows. Given $X_{t-1}$, if $V(X_{t-1}) > d$ then simply choose $X_t \sim P(X_{t-1},\cdot)$. If $V(X_{t-1}) \leq d$, then flip and independent coin with probability of heads $\epsilon$. If the coin comes up heads, choose $X_t \sim Q(\cdot)$ and set $T = t$. If the coin comes up tails, choose $X_t \sim \frac{1}{1-\epsilon}\left(P(X_{t-1},\cdot) - \epsilon\, Q(\cdot)\right)$. Continue in this way for $t = 1, 2, \ldots, T$, i.e. until after the first time the coin comes up heads.

We let the process $\{X'_k\}$ proceed similarly. Each time $V(X'_{t-1}) \leq d$, we flip a new independent coin with probability of heads $\epsilon$. If the coin comes up heads, we set $X'_t = X_T$ and set $T' = t$. If the coin comes up tails, choose $X'_t \sim \frac{1}{1-\epsilon}\left(P(X'_{t-1},\cdot) - \epsilon\, Q(\cdot)\right)$. Continue in this way for $t = 1, 2, \ldots, T'$, i.e. until after the first time the new coin comes up heads.

We thus have by construction that $X_T = X'_{T'}$. To complete the joint definition for times after time $T$ [resp. time $T'$], we let $\{X_{T+m}\}$ and $\{X'_{T'+m}\}$ update identically from $P(\cdot,\cdot)$ so that $X_{T+m} = X'_{T'+m}$ for $m = 1, 2, 3, \ldots$.

We have thus jointly constructed the processes $\{X_k\}$ and $\{X'_k\}$ (each marginally updated according to $P(\cdot, \cdot)$) together with shift-coupling epochs $T$ and $T'$. We now need only bound the tail probabilities $P(\max(T, T') > k)$.

Now, using the drift condition and arguing as in Rosenthal (1995), we have that for any $j > 0$,

$$P(T \geq k) \;\leq\; (1-\epsilon)^{[j]} + P(N_k < j)\,,$$

where $N_k$ is the number of times the process $\{X_m\}$ returns to the set $C$ up to and including time $k$. But then if $r_i$ is the $i$'th return time of $\{X_m\}$ to $C$, then we have

$$P(N_k < j) \;=\; P(r_1 + \ldots + r_j > k) \;=\; P(\lambda^{-(r_1+\ldots+r_j)} > \lambda^{-k})$$

$$\leq\; \lambda^k \mathbf{E}\left(\lambda^{-(r_1+\ldots+r_j)}\right) \;\leq\; \lambda^k \mathbf{E}\left(V(X_0)\right)\,(\lambda^{-1}A)^{j-1}\,,$$

where we have used Markov's inequality and bounds from Rosenthal (1995), and where $A = \sup_{x \in C} \mathbf{E}(V(X_1) \mid X_0 = x) \leq \lambda d + \Lambda$. We thus have that

$$P(T \geq k) \;\leq\; (1-\epsilon)^{[j]} + \lambda^{k-j+1} A^{j-1} \mathbf{E}_\nu(V)\,.$$

Similarly

$$P(T' \geq k) \;\leq\; (1-\epsilon)^{[j]} + \lambda^{k-j+1} A^{j-1} \mathbf{E}_\pi(V)\,.$$

Furthermore it is easily verified that $\mathbf{E}_\pi(V) \leq \frac{\Lambda}{1-\lambda}$. Hence

$$P(\max(T, T') \geq k) \;\leq\; P(T \geq k) \;+\; P(T' \geq k)$$

$$\leq\; 2(1-\epsilon)^{[j]} + \lambda^{k-j+1} A^{j-1} \left(E_\nu(V) + \frac{\Lambda}{1-\lambda}\right)\,.$$

Using Proposition 1, and setting $j = rk + 1$, we thus have

**Theorem 4.** *Suppose a Markov chain $P(\cdot, \cdot)$, on a state space $\mathcal{X}$, with initial distribution $\nu(\cdot)$ and stationary distribution $\pi(\cdot)$, satisfies drift and minorization conditions as in $(i)$ and $(ii)$ above. Then for any $0 < r < 1$, the total variation distance of the ergodic averages to stationarity satisfies*

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) \;-\; \pi(\cdot) \right\| \leq$$

$$\frac{1}{n} \sum_{k=1}^{n} \left(2(1-\epsilon)^{rk} + \lambda^{(1-r)k} A^{rk} \left(E_\nu(V) + \frac{\Lambda}{1-\lambda}\right)\right)\,.$$

**Remarks.**

1. The theorem allows for a range of (sufficiently small) values of $r$ to be used. Naturally, one would usually choose $r$ so as to make the bound as small as possible. An optimal value of $r$ could perhaps be determined by differentiation.
2. By replacing $\sum_{k=1}^{n}$ by $\sum_{k=1}^{\infty}$ in the bound, and using that $A \le \lambda d + \Lambda$, we see that if $r$ is small enough that $\lambda^{(1-r)}(\lambda d+\Lambda)^r < 1$, then

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) \;-\; \pi(\cdot) \right\| \le$$

$$\frac{1}{n}\left(\frac{2(1-\epsilon)^r}{1-(1-\epsilon)^r} \;+\; \frac{\lambda^{(1-r)}(\lambda d+\Lambda)^r}{1-\lambda^{(1-r)}(\lambda d+\Lambda)^r}\left(E_\nu(V)+\frac{\Lambda}{1-\lambda}\right)\right).$$

3. When verifying the minorization condition $(ii)$, it is not necessary to explicitly compute the measure $Q(\cdot)$. Indeed, all that is required is the quantity $\epsilon$, which may be thought of as the "overlap" of the various transition probabilities from different points of $C$.
4. The hypotheses of this theorem are similar to those of recent results of Meyn and Tweedie (1994) and Rosenthal (1995) concerning convergence of individual distributions. However, certain extra conditions are avoided. In Rosenthal (1995) it was required that $d > \frac{2\Lambda}{1-\lambda}$ to allow for coupling at the same time. In Meyn and Tweedie (1994), it was required that the chain have an atom, or be strongly aperiodic, to get complete results and avoid problems related to periodicity. This theorem thus avoids certain troublesome difficulties often associated with the convergence of individual distributions.
5. The bound in the theorem decreases only at rate $O(1/n)$. But it is easily seen that, if $\mathcal{L}(X_0) \neq \pi(\cdot)$, then no faster rate is possible. On the other hand, the distance from stationarity of certain other ergodic averages, such as $\frac{1}{n}\sum_{n+1}^{2n} P(X_k \in \cdot)$, will sometimes decrease at an exponential rate $O(\rho^n)$ for some $\rho < 1$, though it is not clear how such faster bounds can be established using shift-coupling.

One problem with Theorem 4 is that sometimes (e.g. for nearly-periodic chains) the small set $C$ has to be chosen to be extremely small so as to allow $\epsilon$ to be sufficiently large to be of practical value. However, this can lead to $\lambda$ becoming unacceptably large, rendering the bounds of Theorem 4 of little practical use. Thus, we shall now consider an idea for using larger sets, which are subsequently partitioned into a finite number of subsets, each exhibiting a minorization condition. The number of partitioning sets can be arbitrary for the approach adopted here. However since we are usually (at least for MCMC) interested in Markov chains with real eigenvalues which can only have period 1 or 2 (for example reversible chains), and also for computational manageability, we restrict attention to the case of two partitioning sets.

Let $C = \{x \in \mathcal{X} : V(x) \leq d\} = C_1 \cup C_2$, with $C_1$ and $C_2$ disjoint, such that,

(i)(*drift condition*) $\mathbf{E}(V(X_1)|X_0 = x) \leq \lambda V(x) + \Lambda \mathbf{1}_C(x)$ for all $x \in \mathcal{X}$.

(ii) (*minorization condition*) $P(x, \cdot) \geq \epsilon_i Q_i(\cdot)$ for all $x \in C_i$, $i = 1, 2$.

(iii) (*transfer condition*) $Q_1(C_2) \geq \delta$ and $Q_2(C_1) \geq \delta$.

Construct the two processes $X$ and $X'$ as follows. First generate two random variables $Z_1$ and $Z_2$ independently from $Q_1, Q_2$ respectively. Augment $X$ by two indicator processes $I_1$ and $I_2$, defined jointly in the following way. Set $(I_1)_0 = (I_2)_0 = 0$. For $t \geq 1$, if $X_{t-1} \in C_i$ and $(I_i)_{t-1} = 0$, then flip a coin with probability of heads $\epsilon_i$. If a head occurs, set $X_t = Z_i$ and $(I_i)_t = 1$. If a tail occurs, generate $X_t$ independently from $\frac{1}{1-\epsilon_i}(P(X_{t-1}, \cdot) - \epsilon_i Q_i(\cdot))$ and leave $(I_i)_t = 0$. Otherwise generate $X_t$ independently from $P(X_{t-1}, \cdot)$ and leave $(I_i)_t = (I_i)_{t-1}$. Thus, $(I_i)_t = 0$ until such time as there is a regeneration from the set $C_i$, and afterwards $(I_i)_t = 1$.

Construct $X'$ identically (but perhaps with a different starting distribution), and independently (conditional on the values $Z_1$ and $Z_2$, which are the same for both processes); call its corresponding indicator processes $I'_1$ and $I'_2$. The reason for this construction is that we can then define shift-coupling epochs $T_i$ and $T'_i$, for either $i = 1$ or $i = 2$, by $T_i = \min\{t \,|\, (I_i)_t = 1\}$ and $T'_i = \min\{t \,|\, (I'_i)_t = 1\}$. We then clearly have

$$\min\left(\max(T_1, T'_1),\, \max(T_2, T'_2)\right) \;\leq\; \max(S, S')$$

where $S = \inf\{t; (I_1 I_2)_t = 1\}$ and $S' = \inf\{t; (I'_1 + I'_2)_t \geq 1\}$ (i.e. $S$ is the first time when $X$ has regenerated from *both* of $C_1$ and $C_2$, while $S'$ is the first time when $X'$ has regenerated from *either* of $C_1$ and $C_2$). According to the above construction, $S$ and $S'$ are independent.

Exactly as before, we have that $P(S' > k) \leq (1 - \epsilon)^j + \lambda^{k-j+1} A^{j-1} \mathbf{E}_\pi(V)$, where $\epsilon = \min(\epsilon_1, \epsilon_2)$. The corresponding bounds on $S$ are similar but require a few additional observations.

As before, let $N_k$ be the number of returns of the chain $\{X_t\}$ to $C$ up to and including time $k$, and let $r_i$ be the $i^{\text{th}}$ return time. Then we have

**Lemma 5.** *For any non-negative integer $j$,*

$$P(S > k) \leq (1 - \epsilon_1 \epsilon_2 \delta)^j + P(N_{k-1} < j)\,.$$

**Proof.** We use the fact that one way for the process to have regenerated from both $C_1$ and $C_2$ is to regenerate from one, and then immediately from the other. This has probability at least $\epsilon_1\epsilon_2\delta$. Thus, we shall modify the above construction of $X$ for the purpose of this proof, to get a minorization for 2 consecutive iterations of the chain.

We do so by defining new "coin flips" as follows. At each iteration $t$, if $X_t \in C$ and we didn't flip a coin at time $t-1$, then flip one now, with probability $\epsilon_1\epsilon_2\delta$ of a head. If we achieve a head, then generate $X_{t+1}$ from $Q_{i(t)}(C_{2-i(t)}^{-1}Q_{i(t)}(\cdot)$ restricted to $C_{2-i(t)}$, and $X_{t+2}$ from $Q_{2-i(t)}(\cdot)$, where $i(t)=1$ or 2 as $X_t \in C_1$ or $C_2$, respectively. It is easy to verify that this is a joint minorization for the next 2 iterations of the algorithm. The remaining $1-\epsilon_1\epsilon_2\delta$ of probability mass is generated by any independent mechanism.

Then

$$P(S>k) = P(S>k, N_{k-1}<j) + P(S>k, N_{k-1}\ge j)$$

$$= P(S>k, N_{k-1}<j) + P(S>k, N_{k-1}\ge j, \text{ first } j \text{ coin tosses produce tails })$$

$$\le P(N_{k-1}<j) + P(\text{first } j \text{ coin tosses produce tails })$$

$$\le P(N_{k-1}<j) + (1-\epsilon_1\epsilon_2\delta)^j$$

as required. ∎

It remains to control the term $P(N_{k-1}<j)$. Exactly as before, we have that for $\alpha>1$,

$$P(N_{k-1}<j) \le \alpha^{-(k-1)}\mathbf{E}(\Pi_{i=1}^{j}\alpha^{r_i})$$

Now suppose there exists a function $V \ge 1$, and a constant $\alpha>1$, such that

$$\mathbf{E}(V(X_1)|X_0=x) \le \alpha^{-1}V(x)$$

for all $x \in C^c$. Then as in Rosenthal (1995, Lemma 4) we have that $\mathbf{E}(\alpha^{r_1}) \le \mathbf{E}(V(X_0))$, and

$$\mathbf{E}(\alpha^{r_i}|\mathcal{F}_{i-1}) \le \alpha^2 \sup_{x\in C}\mathbf{E}(V(X_2)|X_0=x)\,,$$

where $\mathcal{F}_i$ is the $\sigma$-algebra generated by the chain up to the $i^{\text{th}}$ return to $C$. But

$$\sup_{x\in C}\mathbf{E}(V(X_2)|X_0=x)$$

$$= \sup_{x\in C}\{\mathbf{E}(V(X_2)\mathbf{1}_{C^c}(X_1)|X_0=x) + \mathbf{E}(V(X_2)\mathbf{1}_C(X_1)|X_0=x)\}$$

$$\le \sup_{x\in C}\{\mathbf{E}(V(X_2)\,|\,X_1=x)\} \;+\; \sup_{x\in C}\{\alpha^{-1}\mathbf{E}(V(X_1)\,|\,X_1=x)\}$$

$$\leq (1+\alpha^{-1}) \sup_{x\in C} \mathbf{E}(V(X_1))\,.$$

Hence

$$\mathbf{E}(\alpha^{r_i}|\mathcal{F}_{i-1}) \leq (1+\alpha^{-1})\alpha^2 \sup_{x\in C} \mathbf{E}(V(X_1)|X_0=x).$$

Putting these results together, and setting $A = \sup_{x\in C} \mathbf{E}(V(X_1)\,|\,X_0=x)$ as before, we obtain that

$$P(S>k) \leq (1-\epsilon_1\epsilon_2\delta)^j + A^{j-1}\mathbf{E}(V(X_0))\lambda^{k-2j+1}.$$

Hence by exact analogy to the previous theorem, using that $\mathbf{E}(\max(S,S')) \leq \mathbf{E}(S)+\mathbf{E}(S')$, and setting $j = rk+\frac{1}{2}$, we have

**Theorem 6.** *Suppose a Markov chain $P(\cdot,\cdot)$ on a state space $\mathcal{X}$, with initial distribution $\nu$ and stationary distribution $\pi$, satisfies conditions (i), (ii), and (iii) above. Set $\epsilon = \min(\epsilon_1,\epsilon_2)$. Then for any $0<r<1$ such that $\lambda^{1-2r}A^r<1$,*

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P(X_k\in\cdot) \;-\; \pi(\cdot) \right\| \leq \frac{1}{n}\left(\frac{(1-\epsilon)^r}{1-(1-\epsilon)^r} + \frac{\lambda^{(1-r)}A^r}{1-\lambda^{(1-r)}A^r}\right)(E_\nu(V))$$

$$+\frac{1}{n}\left(\frac{(1-\epsilon_1\epsilon_2\delta)^r}{1-(1-\epsilon_1\epsilon_2\delta)^r} + \frac{\lambda^{(1-2r)}A^r}{1-\lambda^{(1-2r)}A^r}\right)(E_\pi(V))\ ,$$

*where $A = \sup_{x\in C}\mathbf{E}(V(X_1)\,|\,X_0=x) \leq \lambda d+\Lambda$.*

To end this section, we note that the bound in Proposition 1 is bounded above by $\mathbf{E}(T)+\mathbf{E}(T')$. This suggests interest in the expected values of shift-coupling epochs, which is connected to results in potential theory. For example, a paper of Baxter and Chacon (1976) states the following. Assume that our Markov chain *asymptotically* converges weakly to a stationary distribution $\pi$, from any initial distribution. (This assumption can sometimes be verified easily, since there are various general theorems proving asymptotic convergence, cf. Tierney, 1994.) Define the "potential operator" by $G=\sum_{k=0}^{\infty}P^k$, so that, given a measure $\mu$ on $\mathcal{X}$, $\mu G(\cdot) = \sum_{k=0}^{\infty}\int \mu(dx)P^k(x,\cdot)$ is a possibly-infinite measure on $\mathcal{X}$. Then given an initial distribution $\nu$, and an arbitrary probability measure $Q$ on $\mathcal{X}$, Baxter and Chacon's result is that there exists a stopping time $T$ with $\mathcal{L}(X_T)=Q$, if and only if $[(\nu-Q)G]^- = \phi\,d\pi$ is absolutely continuous with respect to $\pi$ (where $[\ldots]^-$ means the negative part of the measure $[\ldots]$). Furthermore, in this case the smallest possible value of $\mathbf{E}(T)$ over all stopping times with $\mathcal{L}(X_T)=Q$, is precisely the essential supremum of $|\phi|$.

This immediately implies information about our shift-coupling epochs, as follows. Given a probability measure $Q$ on $\mathcal{X}$, with $[(\nu-Q)G]^- = \phi\,d\pi$ and $[(\pi-Q)G]^- = \phi'\,d\pi$ as above, define

$$e_Q = \operatorname{ess\,sup}|\phi|\,; \qquad e'_Q = \operatorname{ess\,sup}|\phi'|\,.$$

(For definiteness, set $e_Q$ or $e'_Q$ to $+\infty$ if the negative part of the corresponding measure is not absolutely continuous with respect to $\pi$.) Furthermore, let $M$ be the infimum of the possible values of $\mathbf{E}(\max(T, T'))$, where $T$ and $T'$ are any shift-coupling epochs for our Markov chain $P(\cdot, \cdot)$ starting from the distributions $\nu$ and $\pi$ respectively. Then since we always have $\max(\mathbf{E}(T), \mathbf{E}(T')) \leq \mathbf{E}(\max(T, T')) \leq \mathbf{E}(T) + \mathbf{E}(T')$, we see that

$$\inf_Q \max(e_Q, e'_Q) \;\leq\; M \;\leq\; \inf_Q (e_Q + e'_Q)\,,$$

where $Q$ plays the role of $\mathcal{L}(X_T) = \mathcal{L}(X'_{T'})$. Thus, this theory immediately bounds the quantity M within a factor of 2.

On the other hand, from Proposition 1, the total variation distance of ergodic averages to the stationary distribution $\pi$ is bounded above by $M/n$. We thus conclude

**Proposition 7.** *Let $P(\cdot, \cdot)$ be the transition probabilities for a Markov chain on an arbitrary state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$. Assume that $\lim_{n\to\infty} P^n(x, A) = \pi(A)$ for every $x \in \mathcal{X}$ and for every measurable $A \subseteq \mathcal{X}$. Then given a starting distribution $\nu = \mathcal{L}(X_0)$, we have that for any probability distribution $Q(\cdot)$ on $\mathcal{X}$,*

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) \;-\; \pi(\cdot) \right\| \leq \frac{1}{n}\,(e_Q + e'_Q)\,,$$

*with $e_Q$ and $e'_Q$ as above.*

Quantities like $e_Q$ appear to be very difficult to compute in practice. Thus, it may be difficult to apply this proposition to specific examples.

## 4. EXAMPLES.

We now apply some of the above bounds to a variety of examples of Markov chains with stationary distributions.

### 4.1. A one-dimensional normal example.

We first consider a very simple example, adapted from the bivariate normal example of Schervish and Carlin (1992). We consider the Markov chain defined on the one-dimensional real line by

$$\mathcal{L}(X_k \,|\, X_{k-1} = x) \;=\; N(\frac{x}{2}, \frac{3}{4})\,,$$

where $N(\cdot, \cdot)$ is a normal distribution. Setting $V(x) = 1 + x^2$, it is easily verified that $\mathbf{E}(V(X_1) \,|\, X_0 = x) \;=\; \frac{1}{4}V(x) + \frac{3}{2}$. Hence, choosing $d = 4$, we have that $\mathbf{E}(V(X_1) \,|\, X_0 = x) \;\leq\; \lambda\, V(x)$ whenever $V(x) > d$, with $\lambda = 5/8$. Furthermore it is easily verified (cf. Rosenthal, 1995, Example #1) that for $V(x) \leq d$ we have a minorization condition with

$\epsilon = 0.31$. Hence, taking $\lambda = 5/8$, $\Lambda = 3/4$, and $\epsilon = 0.31$, assuming $\mathbf{E}_\nu(V) = 2$, and choosing $r = 0.18$, we have from Theorem 4 (and Remark 2 following) that

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) \ - \ \pi(\cdot) \right\| \leq \ 50/n\,.$$

### 4.2. A queueing example.

We consider the M/M/1 queue example of Meyn and Tweedie (1994, Section 8). Here $\mathcal{X} = \{0, 1, 2, \ldots\}$, $P(0,0) = p$, $P(0,1) = q = 1-p$, and for $x \geq 1$ we have $P(x, x-1) = p$, $P(x, x+1) = q$. We assume that $p > \frac{1}{2}$, so that this chain has a stationary distribution. Following Meyn and Tweedie, we choose the singleton small set $C = \{0\}$, and choose $V(x) = (p/q)^{x/2} = \pi(x)^{-1/2}$, so that we may take $\lambda = 2\sqrt{pq}$ and $\Lambda = p - \sqrt{pq}$. (We note that the choice $V(x) = \pi(x)^{-1/2}$ is commonly useful; see Roberts and Tweedie, 1994, Theorem 3.3.) Also since $C$ is a singleton, clearly the minorization condition holds with $\epsilon = 1$. Furthermore we may take $d = 1$. The bound of Theorem 4 thus reduces to

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) \ - \ \pi(\cdot) \right\| \leq \ \frac{1}{n}\left(\frac{\lambda^{1-r}(\lambda+\Lambda)^r}{1-\lambda^{1-r}(\lambda+\Lambda)^r}\right)\left(\mathbf{E}_\nu(V) + \frac{\Lambda}{1-\lambda}\right)$$

with $\lambda = 2\sqrt{pq}$ and $\Lambda = p - \sqrt{pq}$. This is minimized by choosing $r = 0$, whence it becomes $\frac{1}{n}\left(\frac{\lambda}{1-\lambda}\right)\left(\mathbf{E}_\nu(V) + \frac{\Lambda}{1-\lambda}\right)$. For example, if $p = 2/3$, then $\lambda = 0.943$ and $\Lambda = 0.195$, so the bound becomes $\frac{1}{n}(16.5)(3.42 + \mathbf{E}_\nu(V))$. If $p = 0.9$, then $\lambda = \Lambda = 0.6$, so the bound becomes $\frac{1}{n}(1.5)(1.5 + \mathbf{E}_\nu(V))$. Hence, in either case, the bound gives very fast convergence of the ergodic averages.

### 4.3. A periodic Markov chain with drift towards 0.

Let us modify the previous example so that $\mathcal{X}$ consists of *all* integers (including the negative ones), and set $P(x, x-1) = p = 1 - P(x, x+1)$ for $x > 0$, and $P(x, x+1) = p = 1 - P(x, x-1)$ for $x \leq 0$. For $p > \frac{1}{2}$, this new model also has a stationary distribution. However, the Markov chain is now periodic (of degree 2), so the individual distributions will in general not converge.

On the other hand, convergence of the ergodic averages follows very similarly to before. Indeed, if we modify $V$ in the obvious way to give $V(x) = (p/q)^{|x|/2}$, then we may keep $C$, $\lambda$, $\epsilon$, and $d$ as in the previous example. The value of $\Lambda$ changes slightly to $\Lambda = \sqrt{p/q} - 2\sqrt{pq}$. We then get sharp bounds on convergence of ergodic averages, exactly as above. If $p = 2/3$, the bound becomes $\frac{1}{n}(16.5)(8.24 + E_\nu(V))$, while if $p = 0.9$ the bound becomes $\frac{1}{n}(1.5)(6 + E_\nu(V))$.

### 4.4. A Metropolis algorithm for a normal density.

Mengersen and Tweedie (1993, Example 4) consider a Metropolis algorithm when the desired stationary distribution $\pi(\cdot)$ is the standard normal distribution $N(0,1)$. They consider a proposal candidate distribution given by $N(x,1)$ when starting at a point $x \in \mathbf{R}$. The Markov chain transitions are then defined as follows. Given that $X_k = x$, we generate a proposal point $y$ from $N(x,1)$, and then either "accept" this proposal and set $X_{k+1} = y$ with probability $\min(1, \phi(y)/\phi(x))$, where $\phi$ is the density function for $\pi(\cdot)$, or "reject" the proposal and set $X_{k+1} = x = X_k$.

Mengersen and Tweedie have shown that this Markov chain satisfies drift and minorization conditions as in Section 3 above. Setting $V(x) = e^{0.48|x|}$ and $C = (-1.15, 1.15)$, they show that we may take $\lambda = 0.95$,

$$\Lambda = \sqrt{2}\Phi(0.48/\sqrt{2})e^{-(0.48)^2/4} + 1 - \frac{1}{\sqrt{2}} - \lambda \doteq 0.188\,,$$

and

$$\epsilon = \frac{1}{\sqrt{2}}\left(\int_{-1.15}^{1.15} e^{-x^2}dx\right) e^{-(1.15)^2} \doteq 0.169\,.$$

They then use these values to apply their general method to the convergence of the individual distributions $P(X_n \in \cdot)$, but the bounds have numerical values on the order of billions of iterations and thus are not very useful in practice.

Using our Theorem 4 above, we can apply these values to get more useful bounds. Indeed, taking $r = 0.06$ and noting that $d = e^{0.48(1.15)} \doteq 1.74$, we obtain that for this example,

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) \;-\; \pi(\cdot) \right\| \le \frac{1}{n}\left(180 + 86\left(4 + \mathbf{E}_\nu(V)\right)\right)\,.$$

These values are quite reasonable, even though they are based on choices of $V$ and $C$ designed to optimize the bounds used by Mengersen and Tweedie.

### 4.5. Antithetic samplers.

Example 6.1 occurs naturally as a Markov chain induced by the Gibbs sampler on a bivariate normal target density with mean 0, variances 1, and correlation $\rho = -1/\sqrt{2}$. $\{X_k\}$ corresponds to one of the one-dimensional components of the chain.

As we stated in the introduction, one of the motivations for the use of shift coupling is to be able to consider "almost periodic" algorithms. Algorithms of this type are sometimes constructed deliberately, in order to create antithetic effects, that subsequently lead to ergodic averages with smaller variances. As a first example of the application of Theorem 4 to antithetic algorithms, consider the following generalization of Example 6.1.

For $-1 < \theta < 1$, define the Markov chain $\{X_k\}$ by

$$\mathcal{L}(X_k | X_{k-1} = x) = N(\theta x, 1 - \theta^2).$$

The stationary distribution for this chain is $N(0,1)$ for each $\theta$. Example 6.1 corresponds to the case $\theta = 1/2$, however for antithetic algorithms, we are more interested in the case where $\theta$ is negative.

Setting $V(x) = 1 + x^2$, it is easy to verify that $\mathbf{E}(V(X_1)|X_0 = x) = \theta^2 V(x) + 2(1-\theta^2)$. Therefore with $d = 4$ as before, $\mathbf{E}(V(X_1)|X_0 = x) \leq \lambda V(x)$ for $X$ such that $V(x) \geq d$, where $\lambda = \frac{1+\theta^2}{2}$. A computation for the minorization measure for $V(x) < d$ gives

$$\epsilon \;=\; \int_{-\infty}^{\infty} \inf_{x \in C} N\left(\theta x, 1 - \theta^2; dy\right) \;=\; 2\Phi\Big(\frac{-|\theta|\sqrt{3}}{\sqrt{1-\theta^2}}\Big),$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function. We can also take $\Lambda = 3(1-\theta^2)/2$.

For a numerical example, taking $\mathbf{E}_\nu(V) = 2$ (say), we apply the last line of Theorem 4 for $|\theta| = 3/4$. We find by trial and error that $r = 0.1$ is nearly optimal among allowable values of $r$; this choice gives that

$$\Big|\Big|\frac{1}{n}\sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot)\Big|\Big| \leq \frac{447}{n}.$$

For $|\theta| = 0.85$, taking $r = 0.08$ as nearly optimal, the bound worsens considerably to $\frac{4969}{n}$. (For $|\theta| = 0.90$, taking $r = 0.06$, the bound becomes $96021/n$, while for $|\theta| = 0.95$ and $r = 0.02$, it is larger than $10^8/n$.) Thus, the bounds give quite reasonable values except when $|\theta|$ is close to 1.

It is interesting to note that the Theorem gives equally good bounds for the rate of convergence for positive and negative values of $\theta$. Therefore to benefit from the antithetic nature of the chain, we need to use an alternative approach, such as that of Theorem 6. We consider this in our final example.

Returning to the MCMC context, Green and Han (1992) suggest the use of antithetic components in a "Gibbs-type" blocking scheme. For example suppose that $\pi$ is a bivariate normal target density with mean $(0,0)$, unit variances and correlation $\rho$. Iteratively construct $(X_n, Y_n)$ from $(X_{n-1}, Y_{n-1})$ as follows.

$$\mathcal{L}(X_n | X_{n-1}, Y_{n-1}) = N(\alpha X_{n-1} + \rho(1-\alpha)Y_{n-1}, (1-\rho^2)(1-\alpha^2))$$

$$\mathcal{L}(Y_n | X_n, Y_{n-1}) = N(\alpha Y_{n-1} + \rho(1-\alpha)X_n, (1-\rho^2)(1-\alpha^2)).$$

It is easy to check that the stationary distribution of this chain is the target distribution $\pi$ above. $\alpha = 0$ corresponds to the Gibbs sampler case, whereas the case where $\alpha$ and $\rho$ have different signs is the antithetic case.

Using the quadratic test function $V(x, y) = 1 + x^2 + y^2$,

$$\mathbf{E}(V(X_n, Y_n)|X_{n-1} = x, Y_{n-1} = y) = 1 + (1 - \rho^2)(1 - \alpha^2)(2 + \rho^2(1 - \alpha)^2)$$

$$+[(\rho^2(1-\alpha)^2 + \alpha)y + \alpha(1-\alpha)\rho x]^2 + (\rho y + \alpha x)^2.$$

We will consider the case $\alpha = -\rho = -1/2$. In this case, by bounding the eigenvalues of the quadratic form above, and a little algebra, we obtain,

$$\mathbf{E}(V(X_n, Y_n)|X_{n-1}, Y_{n-1}) \leq \frac{213}{256} V(X_{n-1}, Y_{n-1}) + \frac{380}{256}$$

which implies that for $V(X_{n-1}, Y_{n-1}) \geq 10$,

$$\mathbf{E}(V(X_n, Y_n)|X_{n-1}, Y_{n-1}) \leq \frac{63}{64} V(X_{n-1}, Y_{n-1}) \, .$$

It follows that we may take $d = 10$, $\lambda = 63/64$, and $\Lambda = 341/256$. A crude bound on the minorization measure for $x, y$ such that $V(x, y) \leq 10$ gives $\epsilon \geq \exp\{\frac{-9\rho^2}{(1-\rho^2)^2}\}$ which equals $e^{-4}$ in the case where $\rho = 1/2$. The bound from Theorem 4 then gives

$$\left|\left|\frac{1}{n}\sum_{k=1}^{n} P\left((X_k, Y_k) \in \cdot\right) - \pi(\cdot)\right|\right| \leq \frac{1}{n}\left(\frac{2}{r}e^4 + \frac{(63/64)^{1-r}98^r}{1-(63/64)^{1-r}98^r}(E_\nu(V) + 88)\right) \, .$$

**4.6. An application of Theorem 6.**

We consider the previous one-dimensional example, with $\mathcal{L}(X_n \mid X_{n-1} = x) = N(\theta x, 1-\theta^2)$. We assume that $-1 < \theta < 0$ and consider applying Theorem 6.

As above, we take $V(x) = 1+x^2$ and $d = 4$, to get $\lambda = \frac{1+\theta^2}{2}$ and $\Lambda = 3(1-\theta^2)/2$. Here $C$ is the interval $[-\sqrt{3}, \sqrt{3}]$ as before. We partition it into $C_1 = [-\sqrt{3}, 0]$ and $C_2 = (0, \sqrt{3}]$. A minorization computation as before gives

$$\epsilon_1 = \epsilon_2 = 2\Phi\left(\frac{-|\theta|\sqrt{3}}{2\sqrt{1-\theta^2}}\right) .$$

The improvement over the previous calculation is the extra factor of 2 in the denominator of the function argument; given the nature of the function $\Phi$ this will represent much more than a factor of 2 increase in the value of $\epsilon$.

To compute the transfer value $\delta$, we recall that by construction

$$\epsilon_1 \, Q_1(C_2) = \int\limits_0^{\sqrt{3}} \inf_{-\sqrt{3} \le x \le 0} N\left(\theta x, 1-\theta^2; dy\right)$$

$$= \int\limits_0^{\frac{\sqrt{3}|\theta|}{2}} N(|\theta|\sqrt{3}, 1-\theta^2; dy) \;+\; \int\limits_{\frac{\sqrt{3}|\theta|}{2}}^{\sqrt{3}} N(0, 1-\theta^2; dy)$$

$$= \Phi\left(\frac{-\sqrt{3}|\theta|}{2\sqrt{1-\theta^2}}\right) - \Phi\left(\frac{-\sqrt{3}|\theta|}{\sqrt{1-\theta^2}}\right) + \Phi\left(\frac{\sqrt{3}}{\sqrt{1-\theta^2}}\right) - \Phi\left(\frac{\sqrt{3}|\theta|}{2\sqrt{1-\theta^2}}\right) .$$

Hence we may take $\delta$ to be this value divided by $\epsilon_1$. We may then directly apply Theorem 6.

As a numerical example, if $\theta = -3/4$, then $\epsilon_1 = \epsilon_2 = 0.667$, and $\delta = 0.493$. Taking $r = 0.1$, and again putting $\mathbf{E}_\pi(V) = \mathbf{E}_\nu(V) = 2$, Theorem 6 thus gives

$$\Big|\Big| \frac{1}{n} \sum_{k=1}^{n} P(X_k \in \cdot) - \pi(\cdot) \Big|\Big| \le \frac{149}{n} .$$

For $\theta = -0.85$, taking $r = 0.05$, the bound becomes $\frac{230}{n}$. This appears to be a substantial improvement over the bounds of the previous example. Furthermore, as $\theta \to -1$ the improvement will become more and more pronounced. We thus conclude that, in this example at least, much can be gained from the multiple-minorization-condition approach of Theorem 6.

**Acknowledgements.** We thank Torgny Lindvall for suggesting the use of shift-coupling, thank David Aldous, John Baxter, and Richard Tweedie for insightful conversations, and thank Alan Gelfand for organizing the conference at Mt. Holyoke College which initiated this collaboration. We thank the referees for many comments. This work was partially supported by EPSRC of the U.K. and by NSERC of Canada.

## REFERENCES


D.J. Aldous (1993), Reversible Markov chains and random walks on graphs. Lecture notes, Dept. of Statistics, University of California, Berkeley.

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

S. Asmussen, P.W. Glynn, and H. Thorisson (1992), Stationarity detection in the initial transient problem. ACM Trans. Modeling Comput. Simulation **2**, 130-157.

P.H. Baxendale (1994), Uniform estimates for geometric ergodicity of recurrent Markov chains. Tech. Rep., Dept. of Mathematics, University of Southern California.

J.R. Baxter and R.V. Chacon (1976), Stopping times for recurrent Markov processes. Illinois J. Math. **20**, 467-475.

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

C. Geyer (1992), Practical Markov chain Monte Carlo. Stat. Sci., Vol. **7**, No. **4**, 473-483.

P.J. Green and X.-L. Han (1992), Metropolis methods, Gaussian proposals, and antithetic variables. In Stochastic Models, Statistical Methods and Algorithms in Image Analysis (P. Barone et al., Eds.). Springer, Berlin.

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

K.L. Mengersen and R.L. Tweedie (1993), Rates of convergence of the Hastings and Metropolis algorithms. Tech. Rep. **93/30**, Dept. of Statistics, Colorado State University.

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

J.G. Propp and B.M. Wilson (1995), Exact sampling with coupled Markov chains and applications to statistical mechanics. Preprint, Dept. Mathematics, Mass. Inst. Technology.

G.O. Roberts and R.L. Tweedie (1994), Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika, to appear.

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

M.J. Schervish and B.P. Carlin (1992), On the convergence of successive substitution sampling. J. Comp. Graph. Stat. **1**, 111–127.

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.

H. Thorisson (1993), Coupling and shift-coupling random sequences. Contemp. Math., Volume **149**.

H. Thorisson (1994), Shift-coupling in continuous time. Prob. Th. Rel. Fields **99**, 477-483.

L. Tierney (1994), Markov chains for exploring posterior distributions. Tech. Rep. **560**, School of Statistics, University of Minnesota. Ann. Stat., to appear.