ELSEVIER Stochastic Processes and their Applications 80 (1999) 211–229

stochastic processes and their applications

Bounds on regeneration times and convergence rates for Markov chains [1]

G.O. Roberts[a,*], R.L. Tweedie[b]

[a] *Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK*
[b] *Department of Statistics, Colorado State University, Fort Collins CO 80523, USA*

Received 10 February 1998; received in revised form 18 September 1998; accepted 25 September 1998

Abstract

In many applications of Markov chains, and especially in Markov chain Monte Carlo algorithms, the rate of convergence of the chain is of critical importance. Most techniques to establish such rates require bounds on the distribution of the random regeneration time T that can be constructed, via splitting techniques, at times of return to a "small set" C satisfying a minorisation condition $P(x,\cdot)\geqslant\varepsilon\varphi(\cdot)$, $x\in C$. Typically, however, it is much easier to get bounds on the time τ_C of return to the small set itself, usually based on a geometric drift function $\mathrm{PV}\leqslant\lambda V+b\mathbb{1}_C$, where $\mathrm{PV}(x)=\mathsf{E}_x(V(X_1))$. We develop a new relationship between T and τ_C, and this gives a bound on the tail of T, based on ε,λ and b, which is a strict improvement on existing results. When evaluating rates of convergence we see that our bound usually gives considerable numerical improvement on previous expressions. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Renewal times; Geometric ergodicity; Rates of convergence; Markov chain Monte Carlo; Shift coupling; Computable bounds

1. Introduction

There has recently been considerable work on finding quantitative bounds on the rate of convergence of Markov chains, either in the finite state case (Diaconis and Stroock, 1991; Jerrum and Sinclair, 1996; Kannan, 1994) or the general state case (Meyn and Tweedie, 1994; Roberts and Rosenthal, 1997; Rosenthal, 1995).

Many of the general methods require a bound on the regeneration times that can be constructed via splitting techniques on returns to "small" sets. However, in practice, one may only have results for the small set return times rather than the regeneration times themselves. Our first results, in Section 2, give relationships between these two. We show in Section 3 that the resulting bounds are strictly better in all cases than previous bounds in Meyn and Tweedie (1994) and Rosenthal (1995).

[1] Work supported in part by NSF Grant DMS 9504561 and EPSRC grant GR/J19900.
* Corresponding author. Tel.: +44-1524-59-3948; fax: +44-1524-59-2681;
E-mail address: g.o.roberts@lancaster.ac.uk. (G.O. Roberts)

0304-4149/99/$ – see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0304-4149(98)00085-4

We then apply these bounds in Sections 4 and 5 to a number of convergence rate methods, showing that considerable improvement results in general from use of the tighter results we give here.

The context in which we work is essentially that of positive Harris recurrent chains: for more details of this structure see (Meyn and Tweedie 1993) (whose notation we follow in general), and Nummelin (1984).

We consider a Markov chain $\boldsymbol{X} = \{X_n,\ n = 0, 1, 2, \ldots\}$ on a general state space X. We will assume that $\boldsymbol{X}$ is π-irreducible for some invariant measure π: that is, when $\pi(A) > 0$, for any $x \in \mathsf{X}$ there exists n such that $P^n(x, A) > 0$, where P^n is the n-step transition law

$$P^n(x, A) = \mathsf{P}(X_n \in A | X_0 = x).$$

We define P_μ and E_μ to be the probability and expectation operators of events conditional on X_0 having the distribution μ, and we write $P^n(\mu, A) = \mathsf{P}_\mu(X_n \in A)$.

For such a chain it is known (Meyn and Tweedie, 1993, Ch. 5) that there always exists some (and indeed many) *small sets* C satisfying, for some constants $k, \varepsilon > 0$ and some probability measure φ, the minorisation condition

$$P^k(x, \cdot) \geqslant \varepsilon\varphi(\cdot), \quad x \in C. \tag{1}$$

In what follows, for simplicity of exposition we shall consider the "strongly aperiodic" case where $k = 1$: that is, we assume that

$$P(x, \cdot) \geqslant \varepsilon\varphi(\cdot), \quad x \in C. \tag{2}$$

The extension to the case $k > 1$ is straightforward as in Rosenthal (1995).

Fundamental to our methodology will be probabilistic coupling techniques (see, for example, Lindvall, 1992). The existence of small sets enables the use of many coupling constructions, similar to those on countable spaces, which at first sight might not seem available on more general spaces. In particular, each time the chain reaches some $x \in C$, there is the possibility of *regeneration*, and standard renewal results can be used when this occurs. In effect, from such an x consider an independent coin toss with probability ε of success: when the toss is successful, move the chain according to the distribution φ, otherwise move with the distribution

$$R(x, \cdot) = (1 - \varepsilon)^{-1}(P(x, \cdot) - \varepsilon\varphi(\cdot)). \tag{3}$$

Overall, the dynamics of the chain are not affected by this coin tossing: however, at each of the times of a success the chain regenerates with the distribution φ independent of x, and so within the sequence of visits to C there is an embedded (random) subsequence of regeneration times.

We will take C to be a specific small set satisfying Eq. (2) and define as in Meyn and Tweedie (1993) $\tau_C = \inf\{n \geqslant 1 : X_n \in C\}$ to be the hitting (or return) time for C, and $\sigma_C = \inf\{n \geqslant 0 : X_n \in C\}$ to be the entry time for C. Note that on C^c these are identical. Our main results give tail properties of the random variable T, defined as the time of first regeneration according to the distribution φ. Specifically, if $X_k \in C$ and the next step has distribution φ then $T = k + 1$ and $X_T \sim \varphi$.

The tail properties of T are couched in terms of ε and the tails of τ_C, or bounds on those tails based on geometric drift conditions.

The splitting construction using φ and R, due to Nummelin (1978) and Athreya and Ney (1978), provides a rapid route to proving limit theorems for Markov chains. In particular, for Harris chains (Meyn and Tweedie, 1993, Ch. 13) it gives simple proofs of well-known results such as

$$||P^n(x,\cdot) - \pi|| \to 0, \quad x \in \mathsf{X}, \tag{4}$$

where $||\mu|| := \sup_A \mu(A)$ is the total variation norm of a signed measure μ of total mass 0. Note that this definition of total variation is that used in, for example, Rosenthal (1995), but it differs by a factor of 2 from that often used in, say, Meyn and Tweedie (1993) or Lindvall (1992). Care needs to be taken in the calculated constants below because of this.

In this paper we are ultimately concerned with the rate of convergence in Eq. (4) in the *geometrically ergodic* case, where for some $M(x) < \infty$ and some $\rho < 1$ we have

$$||P^n(x,\cdot) - \pi|| \leqslant M(x)\rho^n, \quad x \in \mathsf{X}. \tag{5}$$

The problem of identifying ρ or even of getting reasonable bounds on ρ has received much attention in recent years (Meyn and Tweedie, 1994; Baxendale, 1993; Rosenthal, 1995; Roberts and Rosenthal, 1997; Lund and Tweedie, 1996) and appears to be non-trivial in general. It appears to be linked to finding tail properties of T, and this is our main motivation for the results in the next section.

The bounds we find in this paper on ρ in Eq. (5) are strictly better than those in Rosenthal (1995) and Meyn and Tweedie (1994). In some cases this enables substantial numerical improvement in the bounds on Eq. (5). It is notable, however, that in cases where the rate of convergence ρ is very near unity, the results of Rosenthal (1995) are very close to those we find here. Moreover, the constant terms bounding $M(x)$ in Rosenthal (1995) are typically smaller than those we find, so that when $\rho \approx 1$ the numerical bounds on Eq. (5) are not strictly ordered until n becomes quite large.

2. Bounds on generating functions of regeneration times

Let us fix the small set C satisfying Eq. (2), and suppose μ is the initial distribution of X_0. We define the generating functions

$$G_\mu(\beta) = \mathsf{E}_\mu[\beta^{\tau_C}], \quad \hat{G}_\mu(\beta) = \mathsf{E}_\mu[\beta^{\sigma_C}],$$

it is known that $\boldsymbol{X}$ being geometrically ergodic is equivalent to $G_\mu(\beta) < \infty$ for some $\beta > 1$ (that is, to τ_C having geometrically decreasing tails), at least for suitable μ (Meyn and Tweedie, 1993, Ch. 15), including initial distributions concentrated on single points ($\mu = \delta_x$) for π-a.e. x.

Also, define the generating function

$$H_\mu(\beta) = \mathsf{E}_\mu[\beta^{\mathrm{T}}], \quad \beta < \beta^*$$

for the regeneration times; here β^* denotes the radius of convergence of $H_\mu(\beta)$, and our main goal is to find lower bounds on β^*.

There are some existing results in this direction. Meyn and Tweedie (1994) find a simple lower bound $\beta_{\mathrm{MT}} \leqslant \beta^*$, and Rosenthal (1995) derives a more delicate bound

$\beta_{\mathrm{Ros}} \leqslant \beta^*$. We shall find a bound which we show in Section 3 is strictly better than either β_{MT} or β_{Ros}.

We will need the conditional expectation operator

$$\mathsf{E}_x^*[\,\cdot\,] = \mathsf{E}_x[\,\cdot \mid \text{first step according to } R(x,\cdot)],$$

where $R(x,\cdot)$ is as in Eq. (3); and from this we further define

$$G_*(\beta) = \sup_{y\in C} \mathsf{E}_y^*[\beta^{\tau_C}] = \beta \sup_{y\in C} \int R(x,\mathrm{d}y)\mathsf{E}_y[\beta^{\sigma_C}].$$

We then have:

Theorem 2.1. *For β small enough such that*

$$(1-\varepsilon)G_*(\beta) < 1, \tag{6}$$

the generating function of T satisfies

$$H_\mu(\beta) \leqslant \frac{\beta\varepsilon\hat{G}_\mu(\beta)}{1-(1-\varepsilon)G_*(\beta)}. \tag{7}$$

Proof. Let τ_j, $j=1,2,\ldots$ denote increments of times of successive visits to C. At each time $\sigma_0+\tau_1+\cdots+\tau_{j-1}$ of entry to C, we generate outcomes of i.i.d. variables χ_j with

$$\mathsf{P}[\chi_j = 1] = \varepsilon, \quad \mathsf{P}[\chi_j = 0] = 1-\varepsilon.$$

If $\chi_j = 1$ we regenerate using φ; otherwise we choose the next step of $\boldsymbol{X}$ according to $R(x,\cdot)$. Let $B(j) = \{\chi_0 = \chi_1 = \cdots = \chi_{j-2} = 0\}$. We then have, decomposing over the index j of the successful regeneration,

$$\mathsf{P}_\mu[T = n+1] = \sum_{j=1}^{\infty} \mathsf{P}_\mu[\sigma_0+\tau_1+\cdots+\tau_{j-1} = n | B(j)]\varepsilon(1-\varepsilon)^{j-1}. \tag{8}$$

Hence,

$$\begin{aligned} H_\mu(\beta) &= \sum_{n=1}^{\infty} \beta^n \mathsf{P}_\mu[T=n] \\ &= \beta \sum_{j=1}^{\infty}\sum_{n=0}^{\infty} \varepsilon(1-\varepsilon)^{j-1}\beta^n \mathsf{P}_\mu[\sigma_0+\cdots+\tau_{j-1} = n | B(j)] \\ &= \beta \sum_{j=1}^{\infty} \varepsilon(1-\varepsilon)^{j-1} \mathsf{E}_\mu[\beta^{\sigma_0+\cdots+\tau_{j-1}} | B(j)]. \end{aligned}$$

Letting

$$h_j(\beta) = \mathsf{E}_\mu[\beta^{\sigma_0+\cdots+\tau_{j-1}} | B(j)], \quad \mathscr{F}_j = \sigma(X_0,\ldots,X_{\sigma_0+\cdots+\tau_{j-1}},\chi_0,\ldots,\chi_{j-1}),$$

we find

$$\begin{aligned} h_j(\beta) &= \mathsf{E}_\mu[\mathsf{E}[\beta^{\sigma_0+\cdots+\tau_{j-1}}|\mathscr{F}_{j-1}]] \\ &= \mathsf{E}_\mu[\beta^{\sigma_0+\cdots+\tau_{j-2}}\mathsf{E}^*_{X_{\sigma_0+\cdots+\tau_{j-2}}}[\beta^{\tau_C}]] \\ &\leqslant G_*(\beta)h_{j-1}(\beta). \end{aligned}$$

It follows that $h_j(\beta)\leqslant(G_*(\beta))^{j-1}\hat{G}_\mu(\beta)$, and so Eq. (7) holds, at least for $\beta>1$ small enough such that we have Eq. (6). □

Thus, we have that T has tails that decrease geometrically with rate at least β provided β satisfies Eq. (6).

In general, of course, G_* is not easy to find analytically, and as in other approaches to this problem we will consider bounds on G_* using "geometric drift" conditions.

For an arbitrary function f and measure $\nu(\cdot)$ we shall adopt the usual abbreviations

$$\mathrm{PV}(x)=\int_{\mathsf{X}}P(x,\mathrm{d}y)V(y), \qquad \nu(f)=\int_{\mathsf{X}}f(y)\nu(\mathrm{d}y).$$

We suppose that $V:\mathsf{X}\to[1,\infty)$ is a Foster–Lyapunov function satisfying, for our given small set C,

$$\mathrm{PV}\leqslant\lambda V+b\mathbb{1}_C, \tag{9}$$

where $b<\infty$ and $\lambda<1$. We will write

$$\begin{aligned} d &= \sup_{x\in C}V(x), \\ A &= \sup_{x\in C}\int \mathrm{PV}(x)\leqslant\lambda d+b. \end{aligned} \tag{10}$$

In some theoretical calculations it is convenient to assume that $V(x)\equiv 1$ for $x\in C$; obviously, if this does not hold we can change V on C to satisfy this condition without violating Eq. (9), but the new value of b will increase if $d=1$, which turns out to be relevant in some numerical calculations below.

It is standard (Meyn and Tweedie, 1993; Lund and Tweedie, 1996) that for such a V and for all $\beta\leqslant\lambda^{-1}$ we have

$$\begin{aligned} \mathsf{E}_y[\beta^{\tau_C}] &\leqslant V(y), \quad y\notin C, \\ \mathsf{E}_y[\beta^{\tau_C}] &\leqslant \beta[\lambda d+b], \quad y\in C \end{aligned} \tag{11}$$

and so in particular since $V\geqslant 1$ on C

$$\hat{G}_\mu(\lambda^{-1})=\mathsf{E}_\mu[\lambda^{-\sigma_C}]\leqslant\mu(V). \tag{12}$$

Note that for $\mu=\pi$ in particular, we have $\pi(V)\leqslant b/(1-\lambda)$ from Eq. (9).

We now use these and related bounds to find conditions under which Eq. (6) holds.

Theorem 2.2. *For λ as in* Eq. (9), *and any $1\leqslant\beta\leqslant\lambda^{-1}$, define*

$$\phi(\beta)=\log\beta/\log\lambda^{-1} \tag{13}$$

and write

$$J = [A - \varepsilon]/\lambda. \tag{14}$$

Then for all μ *we have*

$$\begin{aligned} \hat{G}_\mu(\beta) &\leqslant \mu(V)^{\phi(\beta)}, \\ G_*(\beta) &\leqslant [J/(1-\varepsilon)]^{\phi(\beta)}. \end{aligned} \tag{15}$$

Proof. For $x \in C$,

$$\mathrm{PV} = ((1-\varepsilon)R + \varepsilon\varphi)V \leqslant A,$$

and so, since $V \geqslant 1$,

$$\mathrm{RV} \leqslant (1-\varepsilon)^{-1}(A-\varepsilon).$$

Using Eq. (11) we have for $x \in C$

$$\begin{aligned} \mathsf{E}_x^*[\lambda^{-\tau_C}] &\leqslant \lambda^{-1}\,\mathrm{RV}(x) \\ &\leqslant J/(1-\varepsilon). \end{aligned}$$

By Jensen's inequality, for $1 \leqslant \beta \leqslant \lambda^{-1}$ this gives as required

$$\begin{aligned} G_*(\beta) &\leqslant [G_*(\lambda^{-1})]^{\phi(\beta)} \\ &\leqslant [J/(1-\varepsilon)]^{\phi(\beta)}. \end{aligned}$$

Arguments for $\hat{G}_\mu(\beta)$ are identical, based on Eq. (12). □

As is clear from the proof above, improvements are possible, for instance, when R has positive mass in C.

We note that Rosenthal states a result (Rosenthal, 1995, Theorem 8) based on a similar construction to our Theorem 2.1. However this contains a small error: he essentially claims that for β small enough such that $(1-\varepsilon)G_C(\beta) < 1$, where $G_C(\beta) = \sup_{y \in C} \mathsf{E}_y[\beta^{\tau_C}]$, one should find $H_\mu(\beta) \leqslant \varepsilon\beta\hat{G}_\mu(\beta)/[1-(1-\varepsilon)G_C(\beta)]$. He then bounds this using a Foster–Lyapunov argument. This is very similar to our Theorem 2.1 with E_μ^* replaced by E_μ.

But this is not possible in general, since one needs to condition on non-successes in the coin-tossing process prior to the first regeneration. To see that Rosenthal's relationship can indeed lead to incorrect bounds, consider a chain on $\mathsf{X} = [0, 101)$. When X_n is in $C = [0, 1)$, either sample from $U(0, 1)$ with probability $\frac{1}{2}$, or move up by 9 otherwise. So C is small, with $\varepsilon = \frac{1}{2}$. When outside C move down deterministically by 1, so once the small set is left, the return time is exactly nine steps. The regeneration time is therefore essentially $9Z$ where Z is a geometric $(\frac{1}{2})$ random variable. But letting $V(x) = x$, we have that $\mathrm{PV} \leqslant 5.5$ for all x, and using Eq. (11) directly leads to a contradiction of Rosenthal (1995), (Theorem 8).

From Theorem 2.2 we immediately derive bounds on β^*, the radius of convergence of H_μ.

Theorem 2.3. *The generating function* $H_\mu(\beta)$ *converges for* $0 < \beta < \beta_{RT}$ *where*
(a) *if* $J < 1$ *then* $\beta_{RT} = \lambda^{-1}$;
(b) *if* $J \geqslant 1$ *then*

$$\log \beta_{RT} = \left(\frac{\log \lambda \log(1-\varepsilon)}{\log J - \log(1-\varepsilon)} \right) \leqslant \log(\lambda^{-1}) \tag{16}$$

provided that $\mu(V) < \infty$.

Proof. (a) If $J < 1$, we have $G_*(\lambda^{-1}) \leqslant (1-\varepsilon)^{-1}$ from Eq. (15), and so β_{RT} is bounded by λ^{-1}, since the numerator of Eq. (7) is also finite for $\beta < \lambda^{-1}$.

(b) If $J \geqslant 1$ then from Eq. (15) we have convergence of the denominator of Eq. (7) for $\beta < \beta_{RT}$ where

$$(1-\varepsilon)[J/(1-\varepsilon)]^{\phi(\beta_{RT})} = 1. \tag{17}$$

Rearranging Eq. (17) we have that β_{RT} solves

$$(1-\varepsilon)^{-(1-\phi(\beta))} = J^{\phi(\beta)},$$

so that

$$\frac{\log \beta_{RT}}{\log \lambda^{-1}} [\log J - \log(1-\varepsilon)] = -\log(1-\varepsilon)$$

giving Eq. (16). Since this gives $\beta_{RT} \leqslant \lambda^{-1}$, the numerator also converges in this region and the result is proved. □

Note that this gives a sufficient condition for the rate of decay of T to be that of the return times to C. If λ is such that $G_x(\lambda^{-1}) < \infty$ for all $x \in C^c$ then $V(x) = \hat{G}_x(\lambda^{-1})$ satisfies Eq. (9) for this λ, $b = \sup_{x \in C} \lambda[G_x(\lambda^{-1}) - 1]$ and $d = 1$. Thus if $\sup_{x \in C} G_x(\lambda^{-1}) < 1 + \varepsilon/\lambda$, we are able to take $\beta_{RT} = \lambda^{-1}$.

We now compare the results in Section 2 to two other related results of which we are aware, given by Meyn and Tweedie (1994) and Rosenthal (1995).

3. Ordering of the bounds on β^*

For ease of comparison, in this section we shall always assume $d = 1$.

Meyn and Tweedie (1994), (Theorem 6.1) consider a chain containing an atom $\alpha \subseteq C$, with the minorisation given explicitly by

$$P(x, \cdot) \geqslant \varepsilon \delta_\alpha(\cdot), \quad x \in C,$$

where δ_α is the atomic measure on α. They show by direct construction that if V solves Eq. (9), then there exists V' satisfying

$$PV' \leqslant \lambda_{MT} V' + b' \mathbb{1}_\alpha, \tag{18}$$

where

$$\lambda_{MT} = [\lambda + b/\varepsilon]/[1 + b/\varepsilon], \quad b' = b + b/\varepsilon. \tag{19}$$

Suppose then that there is at least one solution to Eq. (9); note that if $G_*(\lambda^{-1}) < \infty$, it follows (Meyn and Tweedie, 1993, Ch. 15) that we have at least one such solution.

Since at each time in the atom α we have a regeneration by definition, the time $T = \sigma_\alpha + 1$ is a regeneration time. Thus we are able to use Eq. (18) to obtain (as in Eq. (11)) that $H_\mu(\beta_{\mathrm{MT}}) \leqslant \beta\mu(V')$ also, where

$$\beta_{\mathrm{MT}} = 1/\lambda_{\mathrm{MT}} = [1 + b/\varepsilon]/[\lambda + b/\varepsilon]. \tag{20}$$

This gives β_{MT} as one bound on β^*, which we can compare with the bound β_{RT}.

Using a geometric trials argument more closely related to that in Theorem 2.1, Rosenthal (1995) shows that for any $0 < r < 1$, the tails of T satisfy

$$\mathsf{P}_\mu[T \geqslant n] \leqslant (1-\varepsilon)^{rn} + [\lambda(A/\lambda)^r]^n \mu(V), \tag{21}$$

the main arguments for this are in Rosenthal (1995) although this form is given in Roberts and Rosenthal (1997), (Theorem 6). It follows from Eq. (21) that if

$$[\beta_{\mathrm{Ros}}]^{-1} = \inf_{r \in (0,1)} [(1-\varepsilon)^r \vee \lambda(A/\lambda))^r], \tag{22}$$

then β_{Ros} is also a bound on β^*.

We now evaluate β_{Ros} explicitly, and then show that β_{RT} is the best of the three values for any combination of ε, λ, b.

Theorem 3.1. *For any λ, ε, b satisfying* Eqs. (9) *and (2), we have*

$$\log \beta_{\mathrm{Ros}} = \frac{\log\lambda \log(1-\varepsilon)}{\log J_R - \log(1-\varepsilon)}, \tag{23}$$

where $J_R = A/\lambda$; and the bounds satisfy

$$\beta_{RT} > \beta_{\mathrm{Ros}} > \beta_{\mathrm{MT}}. \tag{24}$$

Proof. From Eq. (9) and the fact that $V \geqslant 1$, we must have $J_R > 1$. Since $(1-\varepsilon)^r$ is decreasing and λJ_R^r is increasing in r, we thus have that $[\beta_{\mathrm{Ros}}]^{-1} = (1-\varepsilon)^r$ for r satisfying

$$(1-\varepsilon)^r = \lambda J_R^r,$$

this gives

$$r = \log\lambda/[\log(1-\varepsilon) - \log J_R] \tag{25}$$

and Eq. (23) follows. Since $J < J_R$ the first inequality in Eq. (24) is thus trivial, since also $\beta_{\mathrm{Ros}} < \lambda^{-1}$.

The relation $\beta_{\mathrm{Ros}} > \beta_{\mathrm{MT}}$ is not as simple to prove. Without loss of generality, in these computations we will assume that $d = 1$ and for further ease of notation we put $\delta = 1 - \lambda$ and note we can write

$$-\log \beta_{\mathrm{MT}} = \log(1 + \lambda\varepsilon/b) - \log(1 + \varepsilon/b) = \log(1 - \delta\varepsilon/(b+\varepsilon)).$$

We thus have

$$\frac{\log \beta_{\mathrm{MT}}}{\log \beta_{\mathrm{Ros}}} = \left[\frac{\log(1 - \delta\varepsilon/(b+\varepsilon))}{\log(1-\delta)}\right]\left[1 - \frac{\log(1 + b/\lambda)}{\log(1-\varepsilon)}\right] \tag{26}$$

and we wish to show that this ratio is less than unity.

We will use the facts that for $y \in (0,1)$, we have $-y^{-1}\log(1-y)$ strictly increasing as y increases, and for $y > 0$, $y^{-1}\log(1+y)$ strictly decreasing as y increases.

The first term on the right of Eq. (26) can thus be written as

$$\left[\frac{-\log(1-\delta\varepsilon/(b+\varepsilon))}{\delta\varepsilon/(b+\varepsilon)}\right]\left[\frac{\varepsilon}{b+\varepsilon}\right]\left[\frac{\delta}{-\log(1-\delta)}\right] < \left[\frac{\varepsilon}{b+\varepsilon}\right] \tag{27}$$

and so the right-hand side of Eq. (26) is less than

$$\left[\frac{\varepsilon}{b+\varepsilon}\right]\left[1 - \frac{-\log(1-\delta\varepsilon/(b+\varepsilon))}{\delta\varepsilon/(b+\varepsilon)}\frac{\delta}{-\log(1-\delta)}\frac{\log(1+b/\lambda)}{\log(1-\varepsilon)}\right]. \tag{28}$$

Now from Eq. (9) we have the inequality $\lambda + b > 1$, that is $b > \delta$; and so we have firstly that $\delta\varepsilon/(b+\varepsilon) \leqslant \varepsilon$ and so

$$\frac{-\log(1-\delta\varepsilon/(b+\varepsilon))}{\delta\varepsilon/(b+\varepsilon)}\frac{1}{-\log(1-\varepsilon)} < \frac{1}{\varepsilon}. \tag{29}$$

Secondly, we have

$$\frac{\log(1+b/\lambda)}{b/\lambda} < \frac{\log(1+(1-\lambda)/\lambda)}{(1-\lambda)/\lambda} = \frac{-\log(1-\delta)}{\delta}\lambda. \tag{30}$$

Inserting Eqs. (29) and (30) into the second term in Eq. (28) shows that the whole expression in Eq. (28) is strictly bounded by $[\varepsilon/(b+\varepsilon)][1+b/\varepsilon] = 1$ as required, and so relationship (24) holds. □

It is clear from the forms of Eqs. (23) and (16) that as $\varepsilon \to 0$, the ratio of the bounds $\beta_{RT}/\beta_{\mathrm{Ros}}$ converges to 1; it is also not hard to prove using methods similar to those above that as $\varepsilon \to 0$ and $\lambda \to 1$ we also have convergence of $\beta_{\mathrm{MT}}/\beta_{\mathrm{Ros}} \to 1$.

We compare these bounds over a wider range by considering a zero-modified variation of the simple random walk on the integer lattice $\mathbb{Z}_+$.

Example 1 (*A reflecting random walk*)**.** We take $P(n,\{n+1\}) = p$ for $n > 0$, $P(n,\{n-1\}) = q = 1 - p > \frac{1}{2}$ for $n > 0$, and $P(0,\{0\}) = \varepsilon$; and to give an appropriate value of b in Eq. (9) in the examples below we need in each case to choose j^* and $P(0,j^*)$, $P(0,j^*+1)$ appropriately.

We fix $C = \{0,1\}$, so that Eq. (2) holds with $\varphi = \delta_0$ when $\varepsilon \leqslant q$. As is well known (cf. Lund and Tweedie, 1996), we can satisfy Eq. (9) with $V(x) = \gamma^x$: the best value of λ achieved is $\lambda^* = \sqrt{4pq}$, at $\gamma^* = \sqrt{q/p}$. In the model where $\varepsilon = q$ and $P(0,\{1\}) = p$, we also get $b = q - \sqrt{pq}$, and in this case $\beta^* = 1/\sqrt{4pq}$ is known to be the exact radius of convergence of $\mathsf{E}_0[\beta^{\tau_0}]$.

However, we are interested in the behaviour of the bounds on β^* over a wider range of b, and in general will not be able to use the stochastic monotonicity results in Lund and Tweedie (1996) to get such exact rates.

Table 1 gives the values of the three bounds above for various such combinations. This is done for a range of values of q, and we see that

Table 1
Bounds on the decay rate of T

q	0.51	0.51	0.51	0.51	0.51	0.51
λ^*	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
ε	0.05	0.25	0.50	0.05	0.25	0.50
b	0.03	0.03	0.03	2.00	2.00	2.00
β_{RT}	1.00033	1.00147	1.00239	1.000009	1.00004	1.00009
β_{Ros}	1.00013	1.00018	1.00019	1.000009	1.00004	1.00008
β_{MT}	1.00013	1.00018	1.00019	1.000005	1.00002	1.00004
q	0.60	0.60	0.60	0.75	0.75	0.75
λ^*	0.9798	0.9798	0.9798	0.8660	0.8660	0.8660
ε	0.05	0.25	0.50	0.05	0.25	0.50
b	0.50	0.50	0.50	1.50	1.50	1.50
β_{RT}	1.0024	1.0115	1.0206	1.0072	1.0357	1.0706
β_{Ros}	1.0023	1.0084	1.0129	1.0070	1.0325	1.0605
β_{MT}	1.0018	1.0068	1.0102	1.0043	1.0195	1.0347

(i) In the top half of the table, where we consider values of λ^* near one, we have two modes of behaviour: for low b, the new bound β_{RT} is clearly better, and the other two bounds $\beta_{\mathrm{Ros}}, \beta_{\mathrm{MT}}$ are very similar; and for larger b this is reversed with the bounds $\beta_{RT}, \beta_{\mathrm{Ros}}$ being similar and β_{MT} clearly inferior.

(ii) In the bottom half of the table, for lower values of λ^*, we see a spread of values of the three bounds, depending on the relation of b, ε; for low ε we have that $\beta_{RT}, \beta_{\mathrm{Ros}}$ are similar but as ε increases in relation to b we have wider relative separation.

With some manipulations we can now get reasonable bounds on the tails of the distribution of T, and we will use these for coupling bounds in evaluating convergence.

4. Shift coupling bounds

Taking an approach based on the use of shift coupling, Roberts and Rosenthal (1997), (Theorem 6) consider the convergence of ergodic averages to the limiting distribution. Instead of the individual terms in Eq. (5) this leads to bounds on sums of the transition probabilities of the form

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P^k(\mu, \cdot) - \pi \right\| \leqslant \frac{1}{n} [\mathsf{E}_\mu[T] + \mathsf{E}_\pi[T]] \tag{31}$$

for any initial measure μ. Using a result essentially as in Eq. (21) they deduce that for any $0 < r < 1$ such that $\lambda^{1-r}[\lambda d + b]^r < 1$, where λ, b, V satisfy Eq. (9),

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P^k(\mu, \cdot) - \pi \right\| \leqslant \frac{1}{n} \left(\frac{2(1-\varepsilon)^r}{1-(1-\varepsilon)^r} + \frac{\lambda^{1-r}[\lambda d + b]^r}{1-\lambda^{1-r}[\lambda d + b]^r} \left(\mu(V) + \frac{b}{(1-\lambda)} \right) \right). \tag{32}$$

Note that, perhaps counterintuitively, there is a numerical benefit to increasing d here and decreasing the calculated value of b, since the terms in $\lambda d + b$ will be unchanged

but the final constant term will be smaller. In what follows, we do not attempt to maximise this effect.

We now develop analogous results using the bounds in previous sections.

Theorem 4.1. (i) *Define* $J_\varepsilon = [1-\varepsilon]^{-1}J$. *Then for any* $1 < \beta < \beta_{RT}$, *and for any initial measure* μ,

$$\mathsf{P}_\mu[T \geqslant n] \leqslant \beta\mu(V)^{\phi(\beta)} \frac{1-\beta(1-\varepsilon)}{1-(1-\varepsilon)J_\varepsilon^{\phi(\beta)}} \beta^{-n} \tag{33}$$

and hence

$$\mathsf{E}_\mu[T] \leqslant \beta\mu(V)^{\phi(\beta)} \frac{1-\beta(1-\varepsilon)}{[1-(1-\varepsilon)J_\varepsilon^{\phi(\beta)}][\beta-1]}. \tag{34}$$

(ii) *Alternatively, for any initial measure* μ

$$\mathsf{E}_\mu[T] \leqslant 1 + [\varepsilon^{-1}\log J_\varepsilon + \log\mu(V)]/\log\lambda^{-1}. \tag{35}$$

Proof. (i) The first result is an improvement on the more obvious inequality

$$\mathsf{P}_\mu[T \geqslant n] \leqslant \beta\mu(V)^{\phi(\beta)} \frac{\varepsilon\beta^{-n}}{1-(1-\varepsilon)J_\varepsilon^{\phi(\beta)}} \tag{36}$$

which follows if we simply apply Markov's inequality to inequality (7) and substitute the results of Theorem 2.2.

To see Eq. (33) refine Markov's inequality as follows. Note that T has a hazard rate $h(n) := P[T = n \mid T > n-1] \leqslant \varepsilon$. Hence,

$$\begin{aligned} H_\mu(\beta) &\geqslant \mathsf{E}_\mu[\beta^{\mathrm{T}} \mid T \geqslant n]\mathsf{P}_\mu[T \geqslant n] \\ &\geqslant \frac{\beta^n\varepsilon}{1-\beta(1-\varepsilon)}\mathsf{P}_\mu[T \geqslant n], \end{aligned}$$

by stochastic comparison. The result follows using the bounds on H_μ in Theorems 2.1 and 2.2.

(ii) To prove Eq. (35), we first use a direct calculation from Eq. (8) to get, for any $\beta < \lambda^{-1}$

$$\begin{aligned} \mathsf{E}_\mu[T] &\leqslant 1 + \mathsf{E}_\mu[\sigma_C] + \varepsilon^{-1}\mathsf{E}^*[\tau_C] \\ &\leqslant 1 + \frac{\mathsf{E}_\mu[\beta^{\sigma_C}-1]}{\beta-1} + \varepsilon^{-1}\frac{\mathsf{E}^*[\beta^{\tau_C}-1]}{\beta-1} \\ &\leqslant 1 + \frac{\mu(V)^{\phi(\beta)}-1}{\beta-1} + \varepsilon^{-1}\frac{J_\varepsilon^{\phi(\beta)}-1}{\beta-1}, \end{aligned} \tag{37}$$

here the second line uses the general inequality $\mathsf{E}[Z] \leqslant (\mathsf{E}[\beta^Z]-1)/(\beta-1)$, and the third uses the bounds in Theorem 2.2.

Now each of the terms on the right in Eq. (35) is monotone decreasing as β decreases in $[1, \lambda^{-1})$. To see this, note that both the relevant terms can be written in the form $[K^{\log\beta}-1]/[\beta-1]$: differentiating and rearranging terms we see that the slope of this function is positive if

$$K^{-\log\beta} + [1-\beta^{-1}]\log K > 1$$

and this holds provided $K \geqslant e^{-1}$, since $\beta > 1$. But now we have that K can be written as $J_\varepsilon^{1/\log\lambda^{-1}}$ or $\mu(V)^{1/\log\lambda^{-1}}$, respectively, and since in each case $J_\varepsilon > 1$ or $\mu(V) > 1$, we have $K > 1$ which is more than is required.

Using this monotonicity, it follows that the best bound in Eq. (37) occurs as $\beta \downarrow 1$, and this limit follows from

$$\frac{d}{d\beta} K^{\log\beta}|_{\beta=1} = \log K,$$

which gives Eq. (35) for each of the terms as stated. □

Substituting these results in Eq. (31), we find:

Theorem 4.2. *For any μ the total variation distance between the ergodic average and the stationary measure can be bounded in two different ways*:

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P^k(\mu,\cdot) - \pi \right\| \leqslant \frac{1}{n} \frac{\beta[1-\beta(1-\varepsilon)]}{[1-(1-\varepsilon)J_\varepsilon^{\phi(\beta)}][\beta-1]} \times (\mu(V)^{\phi(\beta)} + [b/(1-\lambda)]^{\phi(\beta)}) \tag{38}$$

for any $1 < \beta < \beta_{RT}$; and

$$\left\| \frac{1}{n}\sum_{k=1}^{n} P^k(\mu,\cdot) - \pi \right\| \leqslant \frac{1}{n}\{2 + [\log\mu(V) + \log b/(1-\lambda) + (2/\varepsilon)\log J_\varepsilon]/\log\lambda^{-1}\}. \tag{39}$$

We will compare these results for a number of different examples, many of which are becoming standard in the Markov chain Monte Carlo literature, and all of which are described in more detail in Roberts and Rosenthal (1997); we will find that the bounds given by Eqs. (32) and (38) are not strictly ordered, although in all examples bound (39) does improve considerably on both the other bounds in Eqs. (32) and (38).

This is not surprising since it mainly uses approximations relevant to calculating $\mathsf{E}[T]$ rather than $\mathsf{E}[\beta^T]$. It might seem appropriate to calculate the bounds on the ergodic averages using entirely an approach that ignored the geometric rates of convergence, since, for example, we know (Meyn and Tweedie, 1993) that if we have solutions to the (non-geometric) Foster–Lyapunov drift condition

$$PV_0 \leqslant V_0 - 1 + b\mathbb{1}_C,$$

then this gives bounds of the form

$$\mathsf{E}_x[\tau_C] \leqslant V_0(x), \quad x \in C^c.$$

However, in order that we also have finiteness of $\mathsf{E}_\pi[\tau_C]$ it is shown in Meyn and Tweedie (1993), (Ch. 13) that one needs finite second moments on τ_C, and although conditions to ensure this are known (Tuominen and Tweedie, 1994), the finite moment generating functions given by Eq. (9) certainly are one way of approaching this problem for most practical situations.

Example 1 (*Reflecting random walk* (*ctd*))**.** We first examine further the reflecting random walk given in Section 3, under the assumption that $\mu(V) = 2$. Table 2 shows

Table 2
Bounds on ergodic averages for the reflecting random walk

λ^*	0.9798	0.9798	0.8660	0.8660
b	0.5	0.5	1.5	2.0
ε	0.25	0.50	0.50	0.05
β_{RT}	1.0115	1.0206	1.0706	1.0060
β_{Ros}	1.0084	1.0129	1.0605	1.0059
Bound from Eq. (32)	2294	1907	190	823
Bound from Eq. (38)	1094	815	143	1288
Bound from Eq. (39)	395	329	64	367

Table 3
Bounds on ergodic averages for various MCMC models

	Example 2 (a)	Example 2 (b)	Example 3	Example 4 $\lvert\theta\rvert = \frac{3}{4}$
λ	0.625	0.625	0.95	0.7813
b	0.75	1.875	0.188	0.6563
ε	0.31	0.31	0.169	0.0248
d	4	1	1.74	4
β_{RT}	1.095	1.113	1.013	1.0038
β_{Ros}	1.1090	1.104	1.011	1.0038
Bound from Eq. (32)	50	57	498	730
Bound from Eq. (38)	78	70	665	1904
Bound from Eq. (39)	32	29	215	531

four sets of parameter values, and in the first three cases the bounds are ordered with Eq. (39) better than Eq. (38) better than Eq. (32). In the last case Eq. (32) is better than Eq. (38), but again Eq. (39) represents a very substantial improvement.

There seems to be some relationship between the relative values of β_{RT} and β_{Ros} and the bounds found, with Eq. (32) being better than Eq. (38) when β_{RT} and β_{Ros} are close to unity: in this case the difference between the geometric rates is not the leading component of the sums, and the constants in Eq. (32) dominate those in Eq. (38). This is further reflected in the examples below.

However, the main conclusion we come to is that Eq. (39) is a clearcut improvement on the other methods.

Example 2 (*Contracting normals*)**.** Let $P(x,\cdot)$ be the $N(x/2, \frac{3}{4})$ distribution, which has the $\mathrm{N}(0,1)$ distribution as its stationary measure. If we take $V(x) = 1 + x^2$, then as is shown in Roberts and Rosenthal (1997), (Example 6.1) we have $\mathrm{PV} = \frac{1}{4}V + \frac{3}{2}$.

If we take $C = \{y : |y| \leqslant \sqrt{3}\}$, then get the values

$$d = 4, \quad \lambda = \tfrac{5}{8}, \quad b = \tfrac{3}{4}, \quad \varepsilon = 0.31.$$

Alternatively we could take the same set C and the same ε and λ; but if we modify V to be 1 on C then we get $d = 1$, $b = 1.875$.

As shown in Table 3, these choices give different effects. The latter choice actually improves the rate of convergence estimates by about 20%, and improves both the bounds in Eqs. (38) and (39); it does not improve the bound in Eq. (32). If we take

$\mu(V) = 2$, then the best of these bounds is $29/n$, again a clear improvement on the bound from Eq. (32).

Example 3 (*A Metropolis algorithm for the normal density*)**.** Let $\mathsf{X} = \mathbb{R}$ and $P(x, \cdot)$ be the Metropolis algorithm for estimating the $N(0, 1)$ density using $N(x, 1)$ candidates, as described in Rosenthal and Roberts (1997), (Example 6.4) and Mengersen and Tweedie (1996), (Example 4). If we take $V(x) = e^{0.48|x|}$, and $C = (-1.15, 1.15)$ as in Mengersen and Tweedie (1996), then we get the values

$$d = 1.74, \quad \lambda = 0.95, \quad b = 0.188, \quad \varepsilon = 0.169.$$

With $\mu(V) = 2$ again, Table 3 shows that the bounds given by both Eqs. (32) and (38) are larger than Eq. (39), which provides an improvement by a factor of more than 2–3.

Note that in this case we have not attempted to optimise any of the parameter choices, which are based on those in Mengersen and Tweedie (1996). Our value of Eq. (32) is at $r \approx 0.04$: this is better than that of $696/n$ in Roberts and Rosenthal (1997), (Example 6.4) at $r = 0.06$.

Example 4 (*Antithetic samplers*)**.** Finally, as in Roberts and Rosenthal (1997), (Example 6.5), we consider $P(x, \cdot)$ to be a $N(\theta x, 1 - \theta^2)$ distribution, which has as its stationary measure the $N(0, 1)$ distribution for all $-1 < \theta < 1$. Example 2 above corresponds to the case $\theta = \frac{1}{2}$.

Using $V(x) = 1 + x^2$, and $C = \{y : |y| \leqslant \sqrt{3}\}$ as before, then we get the values

$$\lambda = \frac{1 + \theta^2}{2}, \quad b = \frac{3(1 - \theta^2)}{2}, \quad \varepsilon = 2\Phi\left(\frac{-|\theta|\sqrt{3}}{\sqrt{1 - \theta^2}}\right),$$

where Φ is the Normal cumulative distribution function.

Assuming $\mu(V) = 2$ again, Table 3 shows the bounds given for $|\theta| = \frac{3}{4}$. In this case Eq. (32) is noticably better than Eq. (38). The difference appears, by detailed comparison of terms, to be due to the fact that all the terms in Eq. (38) are multiplied by coefficients related to the starting values, whereas the initial term in Eq. (32) is free of these.

Again, Eq. (39) provides considerable improvement over both other approaches. This was also seen with more extreme values of θ, although at values such as $\theta = 0.9$ the constant terms are over 150 000, even for Eq. (39).

5. Applications to total variation convergence bounds

Finally, we turn to the use of the results above in evaluating rates in Eq. (5). The key here is to use the coupling construction leading to Theorems 5 and 12 of Rosenthal (1995).

Observe that if we run two copies of the chain, starting from μ and π, and if we run them independently until they both hit C, then we can throw a single ε-coin: if it is heads we can draw from φ and use this choice for both chains and so they

will couple, and otherwise we can draw for each independently using the appropriate residual distribution for the bivariate chain, and wait for the next return to $C \times C$ to attempt a successful regeneration.

We can then use the regeneration results above together with the coupling bound

$$\| P^n(\mu, \cdot) - \pi \| \leqslant \mathsf{P}_{\mu\times\pi}[T > n] \tag{40}$$

to give rates in Eq. (4) or Eq. (5).

To use the bounds above on $\mathsf{P}_{\mu\times\pi}[T > n]$ we need to control the times of returns of the bivariate process to $C \times C$. Now, if there is a *bivariate* test function for the process described above, we can proceed essentially as noted by Rosenthal (1995). Let $P_2((x, y), \cdot)$ denote the bivariate transition law, and suppose that there exists some function $h : \mathsf{X} \times \mathsf{X} \to [1, \infty)$ such that

$$\int P_2((x, y), (\mathrm{d}w, \mathrm{d}z))h(w, z) \leqslant \lambda_2 h(x, y) + b_2 \mathbb{1}_{C\times C}(x, y) \tag{41}$$

for some $0 < \lambda_2 < 1$ and some $0 < b_2 < \infty$. Then we have from Theorem 4.1 that, defining all quantities in terms of λ_2 and b_2, for any $1 < \beta < \beta_{RT}$

$$\| P^n(\mu, \cdot) - \pi \| \leqslant h_{\mu\times\pi}^{\phi(\beta)} \frac{1 - \beta(1 - \varepsilon)}{1 - (1 - \varepsilon)J_\varepsilon^{\phi(\beta)}} \beta^{-n}. \tag{42}$$

We can compare the behaviour of these bounds with that of Theorem 5 of Rosenthal (1995), which uses Eq. (21) to show that for any $0 < r < 1$,

$$\| P^n(\mu, \cdot) - \pi \| \leqslant (1 - \varepsilon)^{rn} + h_{\mu\times\pi}[\lambda_2(1 + b_2/\lambda_2)^r]^n. \tag{43}$$

In making these comparisons, we need to find the best r in Eq. (43) and the best β in Eq. (42). For the former we will use r as in Eq. (25) to ensure the best asymptotic rate. We now consider the best value of β. This result is couched in terms of a test function V but in the application in this section we will use the bivariate function h in Eq. (41), using λ_2, b_2 in the definition of β_{RT} as required.

Theorem 5.1. *For any V satisfying Eq. (9), let*

$$\xi = \frac{\log \mu(V)}{\log \lambda^{-1}}, \quad \eta = \frac{\log J_\varepsilon}{\log \lambda^{-1}}$$

and write $n' = n - \xi$.

(a) *If $J \geqslant 1$ then for $n' > 0$ bound* (33) *on $\mathsf{P}[T > n]$ is approximately minimised at*

$$\hat{\beta} = \beta_{RT}/[1 + \eta/n']^{1/\eta} > \beta_{RT}\mathrm{e}^{-1/n'}. \tag{44}$$

At this value we have

$$\mathsf{P}_\mu[T > n] < [1 - \hat{\beta}(1 - \varepsilon)] \left(1 + \frac{n'}{\eta}\right) \left(1 + \frac{\eta}{n'}\right)^{n'/\eta} \beta_{RT}^{-n'} \tag{45}$$

which is bounded for all $n' > 0$ by

$$\mathsf{P}_\mu[T > n] < \mathrm{e}\beta_{RT}^{\xi}[1 - \beta_{RT}\mathrm{e}^{-1/n'}(1 - \varepsilon)][1 + n'/\eta]\beta_{RT}^{-n}. \tag{46}$$

(b) *If* $J < 1$ *then for*

$$n' > \eta(1-\varepsilon)/[\lambda^n - (1-\varepsilon)] \tag{47}$$

bound (33) *on* $\mathsf{P}[T > n]$ *is approximately minimised at*

$$\hat{\beta} = \lambda^{-1} = \beta_{RT} \tag{48}$$

and at this value

$$\mathsf{P}_\mu[T > n] < \frac{\mu(V)[1 - \beta_{RT}(1-\varepsilon)]}{1-J}\beta_{RT}^{-n}. \tag{49}$$

Proof. It seems difficult to get a closed form for the best value of β in Eq. (33), so we first find $\hat{\beta}$ to minimise Eq. (36) and then substitute this value in Eq. (33). Numerical results indicate that this is indeed an approximate minimum in the latter case also.

In the notation above, we can write Eq. (36) as

$$\mathsf{P}_\mu[T > n] \leqslant \frac{\varepsilon\beta^{\xi-n}}{1-(1-\varepsilon)\beta^\eta}. \tag{50}$$

This bound holds for all $1 < \beta < \beta_{RT}$.

(a) Assume $J \geqslant 1$. Differentiation shows that Eq. (50) is minimised for fixed $n > \xi$ by choosing $\hat{\beta}$ satisfying

$$(1-(1-\varepsilon)\hat{\beta}^\eta)(n-\xi) = \eta\hat{\beta}^\eta(1-\varepsilon), \tag{51}$$

this gives

$$\hat{\beta}^\eta = \frac{1}{(1-\varepsilon)(1+\eta/(n-\xi))}.$$

Now we observe that $\beta_{RT} = (1-\xi)^{-\eta^{-1}}$, and for $n' > 0$,

$$\left(1+\frac{\eta}{n'}\right)^{n'/\eta} < \mathrm{e}, \tag{52}$$

giving Eq. (44); since $\hat{\beta} < \beta_{RT}$ this minimum is within the range of validity of Eq. (50). Substituting $\hat{\beta}$ in the right-hand side of Eq. (50) gives

$$\varepsilon(1-\varepsilon)^{n'/\eta}\left(1+\frac{\eta}{n'}\right)^{n'/\eta}\left(1+\frac{n'}{\eta}\right),$$

but the right-hand side of Eq. (33) is identical to that of Eq. (50) multiplied by $[1+\hat{\beta}(1-\varepsilon)]/\varepsilon$, and thus the inequality in Eq. (45) follows.

To get bound (46) we again use Eq. (52), and the inequality in Eq. (44).

(b) Suppose instead that $J < 1$. As in Eq. (51), we have that the derivative of the bound in Eq. (50) is negative for all β in the range $1 < \beta < \beta_{RT}$, provided we have n satisfying Eq. (47): and thus the minimum is at $\beta_{RT} = \lambda^{-1}$. Using this in Eq. (34) gives the required result. □

Note that although when $J \geqslant 1$, Eq. (33) only holds for $\beta < \beta_{RT}$, the bound is asymptotically in terms of β_{RT}^{-n} and in Eq. (46) we have given a calculable form the constants and the linear terms in n at this boundary value. From Theorem 3.1, it follows that asymptotically in n the bounds given by Eq. (45) must be better than those from

Eq. (43); however, for finite n the exact relationship will depend on the other leading terms, and as we shall see, in some cases where the values of β_{Ros} and β_{RT} are very close, the linear term $[1+n'/\eta]$ in Eq. (45) may play a role in the comparison.

The use of Eq. (43) or Eq. (42) depends on being able to construct a bivariate test function h, and in general this may be difficult. Before moving on to a comparison in specific models, we also refine the argument leading to Theorem 12 of Rosenthal (1995), which allows us to construct such an h from univariate test functions in many cases, including the examples above.

Theorem 5.2. *Suppose that V satisfies* Eq. (9), *and that $C=\{y:\ V(y)\leqslant d\}$. Consider the bivariate chain X_n, X_n' with transition law P_2. Then there exists a bivariate test function $h:\mathsf{X}\times\mathsf{X}\to[1,\infty)$ satisfying Eq. (41) with*

$$\lambda_2=\lambda+b/2(1+d),\quad b_2=b[1+2d]/[2+2d].$$

Proof. Let $h(x,y)=[V(x)+V(y)]/2$. Then for $(x,y)\notin C\times C$,

$$\begin{aligned}P_2h &\leqslant \lambda h+\frac{b}{2}\mathbb{1}_{[C\times C]^c}\\ &\leqslant \left[\lambda+\frac{b}{2(1+d)}\right]h=\lambda_2 h\end{aligned}$$

and for $(x,y)\in C\times C$,

$$P_2h\leqslant\lambda_2h+b-\frac{b}{2(1+d)}\leqslant\lambda_2h+b_2$$

as claimed. □

For this construction to be useful we require $\lambda_2<1$, which entails

$$d\geqslant\frac{b}{2(1-\lambda)}-1 \tag{53}$$

and this does not always hold, as we see below. By increasing C we may find an appropriate d, but the value of ε may become unacceptable, and the best way to proceed in such a case is not yet known. However, this constraint on d is weaker than that in Rosenthal (1995), (Theorem 12), where the requirement was that $d>2b/(1-\lambda)$.

We can now apply this procedure to the previous examples and consider the differences between Eqs. (45) and (43). Table 4 gives relevant calculations for various examples discussed above. We have assumed throughout that $\mu(V)=2$, so that we have $V_{\mu+\pi}=[2+b/(1-\lambda)]/2$.

In the two variations on Examples 2 and 3 we illustrate the bounds at the values of n that correspond to Eq. (45) taking approximate values of 0.01 and 0.001, respectively. In all cases the ratio of the bounds from Eqs. (43) and (45) is considerable, particularly with Example 3(b) where the difference between $\beta_{\mathrm{Ros}}^{-1}$ and β_{RT}^{-1} makes a substantial difference for moderately large n.

Example 4 illustrates a situation where the constant and linear terms dominate the difference in the geometric terms. Here Eq. (43) gives considerably better bounds than Eq. (45) until n becomes too large to be of practical interest. The situation worsens for values of $|\theta|$ that are larger.

Table 4
Bounds on total variation distances for various MCMC models

	Example 2		Example 3		Example 4	Example 1	Example 1
	(a)	(b)	(a)	(b)	$\lvert\theta\rvert = \frac{3}{4}$	$p = 0.7$ $b = 0.5$	$p = 0.75$ $b = 1.0$
n	45	60	500	650	1000	1000	2000
λ_2	0.7	0.7	0.984	0.984	0.847	0.991	0.991
b_2	0.675	0.675	0.154	0.154	0.591	0.425	0.875
ε	0.31	0.31	0.169	0.169	0.0248	0.5	0.4
d	4	4	1.74	1.74	4	2.33	3
$V_{\mu+\pi}$	2.0	2.0	2.88	2.88	2.5	3.99	4.73
$\beta_{\mathrm{Ros}}^{-1}$	0.881	0.881	0.991	0.991	0.993	0.994	0.996
β_{RT}^{-1}	0.846	0.846	0.983	0.983	0.992	0.990	0.995
Bound (43)	0.011	0.0017	0.046	0.012	0.002	0.018	0.002
Bound (45)	0.010	0.0010	0.010	0.001	0.007	0.005	0.002
Ratio	0.82	0.58	0.21	0.08	3.55	0.29	0.99

Finally, we revisit the geometric Example 1. In the two situations shown in Table 4, we find that the new bounds also are better than those of Eq. (43). However, we have not analysed all of the models in Section 3, because in general they do not obey the conditions of Theorem 5.2. The constraint in Eq. (53) is violated, so that for example, if $p = 0.6$ with $\varepsilon = 0.25$, $b = 0.5$, then we find $d = 1.5$ but Eq. (53) requires $d > 11.4$.

We can overcome this by making the set C larger. This enables us to construct a bivariate h according to the prescription of Theorem 5.2, but if that is done then we have no positive value of ε. This also can be remedied: we do get a minorisation for the k-step chain when we choose $C = \{0, 1, \ldots, k\}$. We can then carry out the k-step calculations appropriate for this case, but the details are very much more cumbersome and we do not carry them out here.

References

Athreya, K.B., Ney, P., 1978. A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493–501.

Baxendale, P.H., 1993. Uniform estimates for geometric ergodicity of recurrent Markov processes. Unpublished report, University of Southern California.

Diaconis, P., Stroock, D., 1991. Geometric bounds for eigenvalues of Markov chains. Ann. App. Probab. 1, 36–61.

Jerrum, M., Sinclair, A., 1996. The Markov chain Monte Carlo method: an approach to approximate counting and integration, 1996. In: Dorit, H. (Ed.), Approximation Algorithms for NP-Hard Problems, PWS, Boston, MA, pp. 482–520.

Kannan, R., 1994. Markov chains and polynomial time algorithms. Proc. 35th Annual IEEE Symp. on Foundations of Computer Science, Computer Society Press, pp. 656–671.

Lindvall, T., 1992. Lectures on the Coupling Method. Wiley, New York.

Lund, R.B., Tweedie, R.L., 1996. Geometric convergence rates for stochastically ordered Markov chains. Math. Oper. Res. 21, 182–194.

Mengersen, K.L., Tweedie, R.L., 1996. Rates of convergence of the Hastings and Metropolis algorithms. Ann. Statistics 24, 101–121.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Meyn, S.P., Tweedie, R.L., 1994. Computable bounds for convergence rates of Markov chains. Ann. Appl. Probab. 4, 981–1011.

Nummelin, E., 1978. A splitting technique for Harris recurrent chains. Z. Wahrscheinlichkeitstheorie und Verw. Geb. 43, 309–318.

Nummelin, E., 1984. General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press, Cambridge.

Roberts, G.O., Rosenthal, J.S., 1997. Shift-coupling and convergence rates of ergodic averages. J. Appl. Probab. 35, 1–11.

Rosenthal, J.S., 1995. Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Statist. Assoc. 90, 558–566.

Tuominen, P., Tweedie, R.L., 1994. Subgeometric rates of convergence of f-ergodic Markov chains. Adv. Appl. Probab. 26, 775–798.