One-Shot Coupling for Certain Stochastic Recursive Sequences

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

(May 2001; revised December 2001; addendum August 2002.)

Abstract. We consider Markov chains $\{\Gamma_n\}$ with transitions of the form $\Gamma_n = f(X_n, Y_n)\,\Gamma_{n-1} + g(X_n, Y_n)$, where $\{X_n\}$ and $\{Y_n\}$ are two independent i.i.d. sequences. For two copies $\{\Gamma_n\}$ and $\{\Gamma'_n\}$ of such a chain, it is well known that $\mathcal{L}(\Gamma_n) - \mathcal{L}(\Gamma'_n) \Rightarrow 0$ provided $\mathbf{E}[\log(f(X_n, Y_n))] < 0$, where $\Rightarrow$ is weak convergence. In this paper, we consider chains for which also $\|\Gamma_n - \Gamma'_n\| \to 0$, where $\|\cdot\|$ is total variation distance. We consider in particular how to obtain sharp *quantitative bounds* on the total variation distance. Our method involves a new coupling construction, one-shot coupling, which waits until time n before attempting to couple. We apply our results to an auto-regressive Gibbs sampler, and to a Markov chain on the means of Dirichlet processes.

Keywords. Markov chain, coupling, convergence bounds, stochastic recursive sequence, one-shot coupling, Gibbs sampler, Dirichlet process.

1. Introduction.

In this paper, we consider Markov chains $\{\Gamma_n\}$ of the form

$$\Gamma_n = f(X_n, Y_n)\,\Gamma_{n-1} + g(X_n, Y_n)\,, \tag{1}$$

where $\{X_n\}$ and $\{Y_n\}$ are two independent i.i.d. sequences. (This fits into the general framework of a *stochastic recursive sequence*; see e.g. Borovkov and Foss, 1992; Propp and

* Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, England. Internet: g.o.robert@lancaster.ac.uk.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: jeff@math.toronto.edu. Supported in part by NSERC of Canada.

Wilson, 1996; Foss and Tweedie, 1998; Diaconis and Freedman, 1999; Jarner and Tweedie, 2000a.)

For two copies $\{\Gamma_n\}$ and $\{\Gamma'_n\}$ of such a chain, it is well known that $\mathcal{L}(\Gamma_n) - \mathcal{L}(\Gamma'_n) \Rightarrow 0$ provided $\mathbf{E}[\log(f(X_n, Y_n))] < 0$, where $\Rightarrow$ is weak convergence (see e.g. Dubins and Freedman, 1966; Elton, 1990; Arnold and Crauel, 1992; Diaconis and Freedman, 1999).

In this paper, we consider chains for which also $\|\Gamma_n - \Gamma'_n\| \to 0$, where $\|\cdot\|$ is total variation distance. (Of course, some conditions are necessary, otherwise e.g. one distribution could be always absolutely continuous while the other is always discrete.) We consider in particular how to obtain *quantitative bounds* on the total variation distance (in the spirit of Meyn and Tweedie, 1994; Rosenthal, 1995; Roberts and Tweedie, 1999).

We present a new coupling construction, which we call *one-shot coupling*, for bounding the total variation distance. We are sometimes able to obtain quantitative bounds on total variation distance which are similar to corresponding quantitative bounds on weak convergence.

We apply our results to two substantive examples. The first (Section 6) is an autoregressive Gibbs sampler, with updates given by

$$\Gamma_n = X_n Y_n \Gamma_{n-1} + Y_n \,,$$

where $\{X_n\}$ and $\{Y_n\}$ are i.i.d. with X_n proportional to a chi-squared random variable, and Y_n an inverse gamma random variable. The second (Section 7) is a Markov chain on the means of Dirichlet processes, given by

$$\Gamma_n \;=\; (1 - Y_n) X_n + Y_n \Gamma_{n-1} \,,$$

where $\{X_n\}$ are i.i.d. $\sim \alpha_0$ for some probability measure α_0 on $\mathbf{R}$, and $\{Y_n\}$ are i.i.d. $\sim \text{Beta}(a, 1)$ where $a > 0$. (This corresponds to a reference measure $\alpha = a\alpha_0$; this process has been studied by Feigin and Tweedie (1989) and Guglielmi and Tweedie (2000) among others.) Each of these examples clearly fits into the framework (1).

For each of these examples, we obtain a bound of the form

$$\|\Gamma_n - \Gamma'_n\| \;\le\; C(\Gamma_0, \Gamma'_0) \; (\mathbf{E}[f(X_1, Y_1)])^n \;. \tag{2}$$

Comparing this with (1), we see that $(\mathbf{E}[f(X_1, Y_1)])^n$ is "essentially" the rate at which $|\Gamma_n - \Gamma'_n| \to 0$ pointwise. That is, we see from (1) that, if we choose $X'_n = X_n$ and $Y'_n = Y_n$, then

$$|\Gamma_n - \Gamma'_n| = |\Gamma_0 - \Gamma'_0| \prod_{i=1}^{n} f(X_i, Y_i)\,.$$

Hence, for large n, by taking logs and using the strong law of large numbers,

$$|\Gamma_n - \Gamma'_n| \approx |\Gamma_0 - \Gamma'_0|\, e^{n\,\mathbf{E}[\log f(X_1, Y_1)]}\,.$$

Now, if log were a linear function, so that $\mathbf{E}[\log f(X_1, Y_1)] = \log \mathbf{E}[f(X_1, Y_1)]$, then we would have $|\Gamma_n - \Gamma'_n| \approx |\Gamma_0 - \Gamma'_0|\ (\mathbf{E}[f(X_1, Y_1)])^n$, which would exactly mimic the total variation distance bound (2). We can therefore say that, bounds of the form (2) (such as the bounds in the examples of Sections 6 and 7 below) "essentially" match the asymptotic pointwise convergence rate, aside from the non-linearity of the log function.

Remark. In fact, in (1), it suffices to replace $f(X_n, Y_n)$ by A_n, and $g(X_n, Y_n)$ by $h(A_n, C_n)$, where where $\{A_n\}$ and $\{C_n\}$ are two independent i.i.d. sequences. (Alternatively, we may instead replace $g(X_n, Y_n)$ by A_n, and $f(X_n, Y_n)$ by $h(A_n, C_n)$.) Indeed, to see this, let $A_n = f(X_n, Y_n)$ and $B_n = g(X_n, Y_n)$, and let C_n be i.i.d. uniform on [0,1], independent of $\{A_m\}$. Then we can replace B_n by $B'_n = h(A_n, C_n)$, where $h(a, c)$ is the inverse c.d.f. of the conditional distribution of B_n given $A_n = a$, evaluated at the point c. In symbols,

$$h(a, c) = F^{-1}_{B_n | A_n = a}(c)$$

This means that $B'_n | A_n$ has the same distribution as $B_n | A_n$. Therefore, the pair (A_n, B_n) has the same distribution as (A_n, B'_n). Hence, can replace B_n by B'_n, which gives the result. However, we do not make use of the result in this paper.

2. A Simple Illustrative Example.

To illustrate the basic idea of one-shot coupling, we present a simple illustrative example, in two variants.

Suppose first that a Markov chain $\{\Gamma_n\}$ is defined simply by

$$\Gamma_n \;=\; \frac{1}{2}\,\Gamma_{n-1}\,. \tag{3}$$

That is, this chain involves no randomness at all. (It corresponds to a special case of (1) for which $f \equiv \frac{1}{2}$ and $g \equiv 0$.) Hence, if $\mathcal{L}(\Gamma_0)$ is a point-mass, then $\mathcal{L}(\Gamma_n)$ will be a point-mass for all n. Furthermore, if $\Gamma_0 = \gamma$ and $\Gamma_0' = \gamma' \neq \gamma$, then clearly $|\Gamma_n' - \Gamma_n| = 2^{-n}|\gamma' - \gamma|$ for all n. Hence $\Gamma_n' - \Gamma_n \to 0$ with probability 1, so also $\Gamma_n' - \Gamma_n \Rightarrow 0$ where $\Rightarrow$ is weak convergence. On the other hand, $\|\Gamma_n' - \Gamma_n\| = 1$ for all n so that $\|\Gamma_n' - \Gamma_n\| \not\to 0$, where $\|\cdots\|$ is total variation distance.

Suppose now that we replace (3) by

$$\Gamma_n \;=\; \frac{1}{2}\,\Gamma_{n-1} \;+\; Y_n\,, \tag{4}$$

where $\{Y_n\}$ are i.i.d. $\sim N(0, \frac{3}{4})$. (This chain is a special case of (1) in which the $\{X_n\}$ are ignored; it was discussed by Schervish and Carlin, 1992; Rosenthal, 1995.) This chain has the stationary distribution $N(0,1)$, to which it converges exponentially quickly in total variation distance. Indeed, if $\Gamma_0 = \gamma$, then $\mathcal{L}(\Gamma_n) = N\left(2^{-n}\gamma,\; 1 - 4^{-n}\right)$. Furthermore, it follows easily from Lemma 1 below that

$$\|N(a,v) - N(b,v)\| \;=\; 1 - 2\Phi(-\frac{1}{2}|b-a|\,/\,\sqrt{v})\,, \tag{5}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds$ is the cumulative distribution function of a standard normal; hence,

$$\|\Gamma_n' - \Gamma_n\| \;=\; 1 - 2\Phi\left(-2^{-n-1}|\gamma' - \gamma|\left(1 - 4^{-n}\right)^{-1/2}\right)\,.$$

In this simple example, we therefore get an exact expression for the total variation distance to stationarity after n steps of the Markov chain. However, if we were unable to do this explicit computation, then how could we construct a coupling to obtain a good bound on $\|\Gamma_n' - \Gamma_n\|$?

The one-shot coupling method would be as follows. We would simultaneously construct $\Gamma_n = \frac{1}{2}\Gamma_{n-1} + Y_n$ and $\Gamma'_n = \frac{1}{2}\Gamma'_{n-1} + Y'_n$, by (a) letting $Y'_m = Y_m$ for $m < n$; and (b) attempting to choose the pair (Y_n, Y'_n) so that $Y'_n = Y_n + \frac{1}{2}(\Gamma_{n-1} - \Gamma'_{n-1})$, to ensure that $\Gamma'_n = \Gamma_n$.

Now, we cannot do step (b) with probability 1 while simultaneously ensuring that $Y_n \sim N(0, \frac{3}{4})$ and $Y'_n \sim N(0, \frac{3}{4})$. Indeed, by (5), step (b) can only be made to succeed with probability $1 - 2\Phi(-\frac{1}{2}\,|\Gamma'_{n-1} - \Gamma_{n-1}|\,/\,\sqrt{3/4})$. Hence, the probability of a successful coupling is given by

$$p \;=\; \mathbf{E}\left[1 - 2\Phi\left(-\frac{1}{2}\,|\Gamma'_{n-1} - \Gamma_{n-1}|\,/\,\sqrt{3/4}\right)\right],$$

where the expectation is taken over the joint distribution of $(\Gamma_{n-1}, \Gamma'_{n-1})$. It then follows from the standard coupling inequality (see e.g. Lindvall, 1992) that $\|\Gamma_n - \Gamma'_n\| \leq 1 - p$.

If the joint distribution of $(\Gamma_{n-1}, \Gamma'_{n-1})$ is known explicitly (as in this simple example), then p can be computed precisely. In a more complicated example (such as those of Sections 6 and 7 below), p would instead be bounded from below. In any case, the one-shot coupling construction provides a method of bounding the total variation distance $\|\Gamma_n - \Gamma'_n\|$ between two copies of the Markov chain.

This one-shot coupling method appears to be more natural and more powerful for this sort of Markov chain (in which $\Gamma'_n - \Gamma_n \Rightarrow 0$), than is the conventional multiple-attempt minorisation coupling considered for example in Rosenthal (1995). Indeed, in that paper, for the chain (4), the best asymptotic convergence rate that could be obtained was 0.964, which is far too high. One-shot coupling avoids the wastage of attempting (and perhaps failing) to couple over and over again. For more about comparing the two methods, see the remark at the end of the next section.

3. One-Shot Coupling.

Consider two different copies $\{\Gamma_n\}$ and $\{\Gamma'_n\}$ of a Markov chain, with the same transition probabilities but with different starting distributions $\mathcal{L}(\Gamma_0)$ and $\mathcal{L}(\Gamma'_0)$. We suppose as in (1) that $\Gamma_n = f(X_n, Y_n)\,\Gamma_{n-1} + g(X_n, Y_n)$ and $\Gamma'_n = f(X'_n, Y'_n)\,\Gamma'_{n-1} + g(X'_n, Y'_n)$, where $\{X_n\}$ and $\{Y_n\}$ are two independent i.i.d. sequences, and where the two collections $\{X_n, Y_n\}$ and $\{X'_n, Y'_n\}$ each have the same pre-specified distribution. However, the joint definition of these two different collections is arbitrary, and may be chosen as convenient to establish convergence properties.

If we simply choose $X'_n = X_n$ and $Y'_n = Y_n$ for all n, and if $\mathbf{E}[\log(f(X_n, Y_n))] < 0$, then it is well-known (and easily seen) that $|\Gamma'_n - \Gamma_n| \to 0$ with probability 1, so that $L(\Gamma_n) - L(\Gamma'_n)) \Rightarrow 0$, where $\Rightarrow$ is weak convergence (see e.g. Billingsley, 1995).

Suppose on the other hand that we wish to bound the total variation distance

$$\|\Gamma_n - \Gamma'_n\| \;\equiv\; \sup_{A \subseteq \mathcal{X}} |\mathbf{P}(\Gamma_n \in A) - \mathbf{P}(\Gamma'_n \in A)|\,.$$

The well-known *coupling inequality* (see e.g. Lindvall, 1992) says that

$$\|\Gamma_n - \Gamma'_n\| \;\leq\; \mathbf{P}(\Gamma_n \neq \Gamma'_n)\,,$$

for any joint distribution of Γ_n and Γ'_n, i.e. for any joint construction of the two collections $\{X_n, Y_n\}$ and $\{X'_n, Y'_n\}$. Hence, our goal shall be to jointly define the two collections $\{X_n, Y_n\}$ and $\{X'_n, Y'_n\}$, in such a way as to make $\mathbf{P}(\Gamma'_n = \Gamma_n)$ as large as possible (for some particular, fixed value of n).

We shall adopt a strategy which we shall call *one-shot coupling*. It may be thought of informally as "don't shoot until you see the whites of their eyes". That is, we shall choose $X'_m = X_m$ and $Y'_m = Y_m$ for $m < n$. It is only on the n'th iteration that we shall attempt to force the two chains to become equal.

On the n^{th} iteration, we shall adopt one of the following two strategies.

1. *X-first.* Choose $X'_n = X_n$. Then, attempt to jointly choose Y_n and Y'_n to make $\Gamma_n = \Gamma'_n$, i.e. to solve the equation

$$f(X_n, Y_n)\,\Gamma_{n-1} + g(X_n, Y_n) = f(X_n, Y'_n)\,\Gamma'_{n-1} + g(X_n, Y'_n)\,.$$

2. *Y-first.* Choose $Y'_n = Y_n$. Then, attempt to jointly choose X_n and X'_n to make $\Gamma_n = \Gamma'_n$, i.e. to solve the equation

$$f(X_n, Y_n)\,\Gamma_{n-1} + g(X_n, Y_n) = f(X'_n, Y_n)\,\Gamma'_{n-1} + g(X'_n, Y_n)\,.$$

We can also express these two strategies symbolically, as follows. Let $h(x, y, \gamma) = f(x, y)\gamma + g(x, y)$, and let $h^{(X)}_{y,\gamma}(x) = h(x, y, \gamma)$ and $h^{(Y)}_{x,\gamma}(y) = h(x, y, \gamma)$. Assume for notational convenience that $h^{(X)}$ and $h^{(Y)}$ are invertible. For our coupling construction, we first choose X_n and Y_n from their appropriate distributions. Then, under the X-first strategy, we set $X'_n = X_n$ and attempt to set $Y'_n = h^{(Y)^{-1}}_{X_n,\Gamma'_{n-1}}(h(X_n, Y_n, \Gamma_n))$. Under the Y-first strategy, we instead set $Y'_n = Y_n$ and attempt to set $X'_n = h^{(X)^{-1}}_{Y_n,\Gamma'_{n-1}}(h(X_n, Y_n, \Gamma_n))$,

Which of the two strategies is better, and how effective it is, will depend on the example considered. (We consider two examples below, one based on an auto-regressive Gibbs sampler, and the other based on Dirichlet process means.) Obviously these two strategies are formally equivalent, and amount to simply re-labeling the X_n as Y_n and vice-versa. Thus, for notational simplicity we focus on the X-first strategy below.

Remark. The one-shot coupling strategy considered in this paper is somewhat related to coupling based on drift and minorisation conditions, as studied previously (e.g. Meyn and Tweedie, 1993, 1994; Rosenthal, 1995; Roberts and Tweedie, 1999). However, in conventional drift/minorisation coupling, the two processes attempt to couple every time they have reached some fixed small set C, and if they fail they seek another opportunity to try again. In our one-shot coupling, the processes merely try to get close to each other, not to some fixed set (for related ideas see Jarner and Tweedie, 2000b), and furthermore they wait until the last possible moment (i.e., the target end time n) before attempting to couple, rather than attempting as often as possible before time n.

4. Probability of successful coupling.

Of course, with either of these two strategies, we are required to overall jointly choose (X'_n, Y'_n) from their correct joint distribution. It will not in general be possible to do this while at the same time *always* assuring that $\Gamma'_n = \Gamma_n$.

To control our probability of success, we use the following general (and standard) lemmas about coupling random variables with densities.

Lemma 1. *Given distributions μ and ν, with densities ξ_1 and ξ_2, it is possible to choose (Z_1, Z_2) such that $Z_1 \sim \mu$, $Z_2 \sim \nu$, and $\mathbf{P}(Z_2 = Z_1) \geq \epsilon$, where*

$$\epsilon \;=\; \int_{\mathcal{X}} \min\left[\xi_1(z),\ \xi_2(z)\right]\, dz\,.$$

Proof. Assume $\epsilon > 0$ (otherwise the result is trivial). Let $Q(\cdot)$ be the probability distribution having density $\epsilon^{-1} \min\left[\xi_1(z),\ \xi_2(z)\right]$. Toss an ϵ-coin. If it comes up heads, choose $W \sim Q(\cdot)$ and set $Z_1 = Z_2 = \epsilon$. If it comes up tails, choose $Z_1 \sim (1-\epsilon)^{-1}\left(\mu(\cdot) - \epsilon\, Q(\cdot)\right)$ $Z_2 \sim (1-\epsilon)^{-1}\left(\nu(\cdot) - \epsilon\, Q(\cdot)\right)$, conditionally independently. Then it is easily verified that overall $Z_1 \sim \mu(\cdot)$ and $Z_2 \sim \nu(\cdot)$, and furthermore $\mathbf{P}(Z_2 = Z_1) \geq \epsilon$, as claimed. ∎

Using this lemma together with the usual change-of-variable theorem, and replacing Z_1 with $\zeta(Z_1)$, we obtain

Lemma 2. *Given distributions μ and ν, with densities ξ_1 and ξ_2, and a C^1 one-to-one function ζ, it is possible to choose (Z_1, Z_2) such that $Z_1 \sim \mu$, $Z_2 \sim \nu$, and $\mathbf{P}(Z_2 = \zeta(Z_1)) = \epsilon$, where*

$$\epsilon \;=\; \int_{\mathcal{X}} \min\left[\xi_1(z),\ \xi_2(\zeta(z))\ |\zeta'(z)|\right]\, dz\,.$$

Now, for the X-first strategy (say), suppose it is known that $X_n = x$ and $\Gamma_{n-1} = \gamma$ and $\Gamma'_{n-1} = \gamma'$. If Y_n has density ξ, and ϕ is differentiable where $\phi(y) \equiv \phi_{x,\gamma,\gamma'}(y) = {h^{(Y)}}^{-1}_{x,\gamma'}\left(h(x,y,\gamma)\right))$, then (by Lemma 2) the probability of success of the X-first strategy at time n is equal to

$$\int_{\mathcal{X}} \min\left[\xi(y),\ \xi(\phi_{x,\gamma,\gamma'}(y)) J_{x,\gamma,\gamma'}(y) |\phi'_{x,\gamma,\gamma'}(y)|\right]\, dy\,.$$

We thus obtain

Theorem 3. *If Y_n has density ξ, and $\phi_{x,\gamma,\gamma'}$ is differentiable, then the probability of successful coupling at time n is equal to*

$$\int_{\mathcal{X}} \min\left[\xi(y),\ \xi(\phi_{x,\gamma,\gamma'}(y))\ |\phi'_{x,\gamma,\gamma'}(y))|\right]\ dy\,.$$

5. Delayed one-shot coupling.

For the above coupling construction (with the X-first strategy, say), it is seen that the probability of success of the coupling when choosing Y_n and Y'_n depends on the just-chosen value of X_n (which was chosen to be equal to X'_n).

To deal with this, we shall sometimes use a strategy of *delayed one-shot coupling* as follows. After choosing $X'_m = X_m$, we shall check if a certain condition C_m is satisfied. (Here C_m may depend on X_m, Γ_m, Γ'_m.) If C_m is satisfied, we go ahead and attempt to use Y_m and Y'_m to couple, as above. If it is not, then we instead choose $Y'_m = Y_m$, and wait until time $m+1$ to again attempt to couple.

Suppose we begin our attempted coupling at time n, and allow up to c chances to attempt to couple. Suppose that for all $m \geq n$, the probability that an attempted coupling succeeds at time m, given that C_m has occurred, is at least ϵ. Suppose further that the probability of C_m is at least δ for all $m \geq n$, even conditional on any past failed coupling attempts. Then at time m, with probability at least $1-(1-\delta)^c$ we will have a chance to attempt to couple. Hence, there will be probability at least $\epsilon[1-(1-\delta)^c]$ of successfully coupling by time $n+c$.

We thus obtain

Theorem 4. *Let $n \in \mathbf{N}$. Suppose $\mathbf{P}(C_m \,|\, \mathcal{F}_{m-1}) \geq \delta$ for all $m \geq n$, where $\mathcal{F}_i = \sigma(\Gamma_0, \Gamma'_0, X_0, \ldots, X_i, Y_0, \ldots, Y_i)$. Suppose further that*

$$\mathbf{P}(\text{couple at time } m \,|\, C_m, \mathcal{F}_{m-1}) \geq \epsilon\,, \qquad m \geq n\,.$$

Then with the above delayed one-shot coupling scheme,

$$\mathbf{P}(\text{couple by time } n+c) \geq \epsilon[1-(1-\delta)^c]\,.$$

6. A Gibbs Sampler Example.

Suppose that $Y_1, \ldots, Y_J \sim N(\mu, \tau^{-1})$, where $\{Y_i\}_{i=1}^J$ are data ($J \geq 2$) and where μ and τ are unknown, Assume μ and τ have flat priors on $\mathbf{R}$ and $\mathbf{R}^+$, respectively. Consider running a Gibbs sampler on the pair (μ, τ). Then the updates for μ are Normal, and the updates for τ are Gamma.

Given a sequence of i.i.d. Gamma($(J+2)/2$, 1) random variables, $\{G_i,\ i = 1, 2 \ldots\}$, and independent i.i.d. standard normal variables $\{N_i,\ i = 1, 2, \ldots\}$, we can implement the algorithm according to the following recursions. Given μ_t and τ_t,

1. set $\mu_{t+1} = \bar{y} + N_{i+1}/(J\tau_t)^{1/2} \sim \pi(\mu|\tau_t)$;
2. set $\tau_{t+1} = G_{t+1} \times \left(S/2 + (J/2)(\bar{y} - \mu_{t+1})^2\right)^{-1} \sim \pi(\tau|\mu_{t+1})$.

By combining these two updates, we see that τ_t^{-1} is a Markov chain which iterates following a form of random auto-regression:

$$\tau_{t+1}^{-1} = [N_{t+1}^2/(2G_{t+1})]\tau_t^{-1} + [S/(2G_{t+1})]\,. \tag{6}$$

We can write this equation as

$$\Gamma_t = X_t Y_t \Gamma_{t-1} + Y_t\,,$$

where $\Gamma_t = \tau_t^{-1} = \sigma_t^2$, $X = N_t^2/S$, and $Y_t = S/2G_t$. Here $\Gamma_t \geq 0$ and $X_t \geq 0$.

Now, to get $\Gamma_t = \Gamma_t'$, we need

$$X_t' Y_t' \Gamma_{t-1}' + Y_t' = X_t Y_t \Gamma_{t-1} + Y_t\,.$$

We adopt the X-first strategy of Section 3. That is, we choose $X_t' = X_t$, and attempt to choose Y_t and Y_t' to make $\Gamma_t' = \Gamma_t$.

To proceed, we set $R_t = Y_t'/Y_t$ and $D_t = |\Gamma_{t-1}' - \Gamma_{t-1}|$, so we need

$$X_t R_t \Gamma_{t-1}' + R_t = X_t \Gamma_{t-1} + 1\,,$$

or

$$R_t = \frac{X_t \Gamma_{t-1} + 1}{X_t \Gamma_{t-1}' + 1}\,.$$

It follows that

$$|R_t - 1| = \frac{X_t|\Gamma'_{t-1} - \Gamma_{t-1}|}{1 + X_t\Gamma'_{t-1}} \le X_t D_t\,,$$

where we have used that $X_t \ge 0$ and $\Gamma'_{t-1} \ge 0$.

Furthermore, X_t and D_t are *independent*. Hence, the probability of *not* coupling by time t is at most

$$\mathbf{E}\left[L(R_t - 1)\right] \le \mathbf{E}\left[L_* X_t D_t\right] = L_* \mathbf{E}[X_t]\mathbf{E}[D_t]\,,$$

where $L(\epsilon)$ is the total variation distance between Z and $(1+\epsilon)Z$ when $Z \sim \Gamma(\frac{J}{2}+1, 1)$, and $L_* = \sup_{0<|\epsilon|<\infty} L(\epsilon)/|\epsilon|$.

Here $\mathbf{E}[X_t] = 1$ is easy to compute. Also D_t is $|\Gamma_0 - \Gamma'_0|$ times a product of previous N's and G's, so it isn't hard to compute either. Indeed, $\mathbf{E}[D_t] = \mathbf{E}[N^2]^t\mathbf{E}[\frac{1}{2G}]^t|\Gamma'_0 - \Gamma_0| = 1^t(1/J)^t = J^{-t}|\Gamma'_0 - \Gamma_0|$.

As for L_*:

Lemma 5. $L_* \le \frac{J}{2} + 1$.

Proof. From Lemma 2, the probability of successfully coupling is equal to

$$\int_0^\infty \min\left[\frac{x^{J/2}e^{-x}}{\Gamma(\frac{J}{2}+1)}\,,\ \frac{((1+\epsilon)x)^{J/2}e^{-(1+\epsilon)x}}{\Gamma(\frac{J}{2}+1)}(1+\epsilon)\right] dx$$

$$= \frac{1}{\Gamma(\frac{J}{2}+1)} \int_0^\infty x^{J/2}e^{-x} \min\left[1,\ (1+\epsilon)^{\frac{J}{2}+1}e^{-\epsilon x}\right] dx\,.$$

However $L(\epsilon)$ is simply one minus this probability, so that

$$L(\epsilon) = 1 - \frac{1}{\Gamma(\frac{J}{2}+1)} \int_0^\infty x^{J/2}e^{-x} \min\left[1,\ (1+\epsilon)^{\frac{J}{2}+1}e^{-\epsilon x}\right] dx$$

$$= \frac{1}{\Gamma(\frac{J}{2}+1)} \int_0^\infty x^{J/2}e^{-x} \max\left[0,\ 1-(1+\epsilon)^{\frac{J}{2}+1}e^{-\epsilon x}\right] dx\,.$$

since $\frac{1}{\Gamma(\frac{J}{2}+1)}\int_0^\infty x^{J/2}e^{-x}dx = 1$. Furthermore, for $\epsilon > 0$, the above "max" changes from equaling its first argument (i.e. 0) to equaling its second argument, precisely at the point $x = \epsilon^{-1}(\frac{J}{2}+1)\log(1+\epsilon)$. Hence, for $\epsilon > 0$,

$$L(\epsilon) = \frac{1}{\Gamma(\frac{J}{2}+1)} \int_{\epsilon^{-1}(\frac{J}{2}+1)\log(1+\epsilon)}^\infty x^{J/2}e^{-x}\left(1 - (1+\epsilon)^{\frac{J}{2}+1}e^{-\epsilon x}\right) dx\,. \qquad (7)$$

Now, let $d(s) = L(e^s - 1)$, and let $\beta = e^s$. By the invariance of total variation distance under monotone transformations, the distance between $\Gamma(\frac{J}{2}+1, 1)$ and e^s times a $\Gamma(\frac{J}{2}+1, 1)$ random variable, is equal to the distance between e^s times a $\Gamma(\frac{J}{2}+1, 1)$ random variable and e^{2s} times a $\Gamma(\frac{J}{2}+1, 1)$ random variable. Since total variation distance is a metric, it follows from the triangle inequality that $d(2s) \leq 2d(s)$ for all $s \geq 0$. Thus since d is differentiable at 0, for all $s \geq 0$, $d(s) \leq sd'(0)$. Note also that $d'(0) = L'(0)$. Thus $L(e^s - 1) \leq sL'(0)$, that is $L(\epsilon) \leq \log(1+\epsilon)L'(0)$ which is in turn less than $\epsilon L'(0)$. Therefore $L(\epsilon) \leq \epsilon L'(0)$ for all $\epsilon \geq 0$, and thus $L^* \leq L'(0)$.

We compute the value of $L'(0)$ explicitly, from (7). The contribution to the derivative from the fact that the limits of integration vary with ϵ is 0, since the integrand converges to 0 at the limits of integration. Also $L(0) = 0$, and by L'Hôpital's Rule $\lim_{\epsilon \to 0} \epsilon^{-1} \log(1+\epsilon) = 1$. Hence,

$$L'(0) = \lim_{\epsilon \to 0} \frac{1}{\Gamma(\frac{J}{2}+1)} \int_{\epsilon^{-1}(\frac{J}{2}+1)\log(1+\epsilon)}^{\infty} x^{J/2} e^{-x} \left(x e^{-\epsilon x} (1+\epsilon)^{J/2+1} - (J/2+1)(1+\epsilon)^{J/2} \right) dx$$

$$= \frac{1}{\Gamma(\frac{J}{2}+1)} \int_{\frac{J}{2}+1}^{\infty} x^{J/2} e^{-x} (x - (\frac{J}{2}+1))\, dx \,.$$

$$\leq \frac{1}{\Gamma(\frac{J}{2}+1)} \int_{0}^{\infty} x^{J/2+1} e^{-x}\, dx$$

$$= \frac{\Gamma(\frac{J}{2}+2)}{\Gamma(\frac{J}{2}+1)} = \frac{J}{2} + 1\,,$$

which gives the result. ∎

From this lemma, it follows that

$$\mathbf{P}(\text{not coupling by time } n) \;\leq\; (\frac{J}{2}+1)\, J^{-n} |\Gamma_0' - \Gamma_0|\,.$$

Hence, we have

Theorem 6. *For two copies $\{\tau_n\}$ and $\{\tau'_n\}$ of the auto-regressive Gibbs sampler algorithm given by (6), the total variation distance between them at time n satisfies*

$$\|\tau_n - \tau'_n\| \;\leq\; (\frac{J}{2} + 1)\, J^{-n} \left| \frac{1}{\tau_0} - \frac{1}{\tau'_0} \right| .$$

This theorem gives an asymptotic convergence rate of J^{-n}. Note that in (6), we have $\mathbf{E}[N_{t+1}^2/(2G_{t+1})] = 1/J$. Hence, the asymptotic rate in the theorem is essentially (aside from the non-linearity of the log function) the same rate that $|\Gamma'_n - \Gamma_n|$ goes pointwise to 0; see the discussion at the end of the Introduction.

7. An example with Dirichlet process means.

Feigin and Tweedie (1989) and Guglielmi and Tweedie (2000) consider the following Markov chain on the means of dirichlet processes:

$$\Gamma_n \;=\; (1 - Y_n)X_n + Y_n\Gamma_{n-1}\,, \qquad n = 1, 2, 3, \ldots\,, \tag{8}$$

where $\{X_n\}$ are i.i.d. $\sim \alpha_0$ for some probability measure α_0 on $\mathbf{R}$, and $\{Y_n\}$ are i.i.d. $\sim$ Beta$(a, 1)$ where $a > 0$. (This corresponds to a reference measure $\alpha = a\alpha_0$ for the Dirichlet process.)

Guglielmi and Tweedie (2000) present a detailed study in which they apply the general theory of quantitative convergence rates for Markov chains (Meyn and Tweedie, 1994; Rosenthal, 1995; Roberts and Tweedie 1999, 2000) to obtain precise quantitative upper bounds (their Theorems 2 and 3) on the total variation distance of this process to stationarity after n steps. These theorems are quite impressive. However, except in a very special case (their Theorem 2(i)), the resulting bounds appear to be overly conservative numerically.

We note that it is necessary to assume that α_0 is non-degenerate. Indeed, suppose instead that α_0 is a point-mass at some $a \in \mathbf{R}$. Suppose further that initially we choose $\Gamma_0 < a < \Gamma'_0$. Then we will have $\Gamma_n < a < \Gamma'_n$ for all n. In this case, we clearly do not get convergence at all in total variation distance. (The Guglielmi-Tweedie paper implicitly avoids such degenerate α_0, by having $L < U$ in their equation (1), and by having $\delta > 0$ in their equation (15).) We thus assume from now on that α is non-degenerate.

We proceed here to obtain quantitative bounds for this chain which are more direct and sharp than are those of Guglielmi and Tweedie (2000).

Iterating equation (8), we see that

$$\Gamma_n \;=\; (Y_1 \dots Y_n)\,\Gamma_0 + R_n\,,$$

where R_n is a (complicated) random variable which does not depend on Γ_0. Furthermore, recall that $0 \le Y_i \le 1$. It follows that, if we run two copies of the chain $\{\Gamma_n\}$ and $\{\Gamma'_n\}$, using the same values of $X_1, \dots, X_n, Y_1, \dots, Y_n$ for each chain, then

$$|\Gamma_n - \Gamma'_n| = (Y_1 \dots Y_n)|\Gamma_0 - \Gamma'_0| \to 0\,, \qquad \text{with probability } 1\,.$$

since the $\{Y_i\}$ are i.i.d. $\sim \text{Beta}(a, 1)$ which is concentrated on $(0, 1)$.

Now let us consider the X-first strategy of Section 3. That is, we consider letting $X_1, \dots, X_n$ and $Y_1, \dots, Y_{n-1}$ be the same for both processes, but attempting to choose Y_n and Y'_n dependently so that

$$(1 - Y_n)X_n + Y_n\Gamma_{n-1} \;=\; (1 - Y'_n)X_n + Y'_n\Gamma'_{n-1}\,,$$

i.e.

$$Y'_n = Y_n \left(\frac{\Gamma_{n-1} - X_n}{\Gamma'_{n-1} - X_n} \right) \equiv R_n Y_n\,, \tag{9}$$

where

$$R_n = \frac{\Gamma_{n-1} - X_n}{\Gamma'_{n-1} - X_n}\,.$$

(If instead we used the Y-first strategy, by fixing Y_n and varying X_n and X'_n, we would instead need

$$X'_n = X_n + \frac{Y_n}{1 - Y_n}(\Gamma_{n-1} - \Gamma'_{n-1})\,,$$

but this is difficult to work with, since $X_n \sim \alpha_0$ and α_0 is essentially arbitrary.)

By Lemma 2, we can jointly choose (Y_n, Y'_n) to satisfy (9), while simultaneously ensuring that $Y_n, Y'_n \sim \text{Beta}(a, 1)$, with probability

$$\epsilon_* \;=\; \int \min\left[f(s),\; R_n\, f(s\, R_n)\right]\, ds\,,$$

where $f(s) \propto s^{a-1}$ is the density of the Beta$(a,1)$ distribution.

The problem is that the "scaling factor" $R_n = \frac{\Gamma_{n-1}-X_n}{\Gamma'_{n-1}-X_n}$ in the above expression is multiplicative, and furthermore depends heavily on the distribution $\alpha_0(\cdot)$ of X_n, so it's not clear how we can control this. To proceed, we let Q_r be the total variation distance between a Beta$(a,1)$ random variable, and r times a Beta$(a,1)$ random variable. We have the following.

Lemma 7. *$Q_r = 1-r^a$ if $r \le 1$, and $Q_r = 1-r^{-a}$ if $r \ge 1$. In either case, $Q_r \le a|1-r|$.*

Proof. Let $Z \sim \text{Beta}(a,1)$, and $Z' = rZ$. Let $W = \log Z$ and $W' = \log Z'$. Since log is a one-to-one function, $\|W' - W\| = \|Z' - Z\| = Q(r)$. Now, we compute that W and W' have densities ae^{ax} $(x<0)$ and $ae^{ax-\beta}$ $(x < \beta/a)$ (where $\beta = \log r$), respectively. We then compute that the total variation distance between the log variables is equal to $1 - e^{-|\beta|a}$. Thus, $Q_r = 1-r^a$ if $r \le 1$, and $Q_r = 1-r^{-a}$ if $r \ge 1$.

The final inequality follows by noting that $1-(1-x)^a \le ax$ for $a \ge 1$, $x \ge 0$, and setting $x = 1-r$ (and noting that $1 - \frac{1}{r} \le r-1$ if $r \ge 1$). ∎

Let $k_A = \inf_{g \in \mathbf{R}} \mathbf{P}(|X-g| \ge A)$ for $A > 0$, where $X \sim \alpha_0$. We assume that $\limsup_{A \searrow 0} k_A > 0$; this certainly follows if α_0 has density with respect to Lebesgue measure which is bounded by K, for then $k_A = \inf_{g \in \mathbf{R}} \mathbf{P}(|X-g| \ge A) \ge 1 - 2AK$.

To proceed, we shall again adopt the X-first strategy of Section 3, but this time with the delayed coupling modification of Section 5. That is, we first choose $Y_1, \ldots, Y_{n-1}$ and $X_1, \ldots, X_n$ (with $Y'_i = Y_i$ and $X'_i = X_i$ for all appropriate i). We then compute

$$R_n \;=\; \frac{\Gamma_{n-1} - X_n}{\Gamma'_{n-1} - X_n}\,.$$

If $|R_n - 1| \le \delta$ we then attempt to choose Y_n and Y'_n to make $\Gamma'_n = \Gamma_n$, as in (9). This succeeds with probability Q_{R_n}. On the other hand, if $|R_n - 1| > \delta$, then we choose Y_n and Y'_n independently, and proceed to time $n+1$.

We now define

$$L^{\text{Beta}}(\epsilon) = (1 - \|\text{Beta}(a,1) - [\epsilon + \text{Beta}(a,1)]\|)\,, \quad \text{and} \quad L^{\text{Beta}}_* = \sup_{\epsilon > 0} \epsilon^{-1} L^{\text{Beta}}(\epsilon)\,.$$

Lemma 8. $L_*^{\mathrm{Beta}} = a$.

Proof. We have that $L^{\mathrm{Beta}}(\epsilon) = 1 - (1-\epsilon)^a$ at least for $0 \le \epsilon \le 1$ and $a \ge 1$. Since this function is concave,

$$\sup_{\epsilon>0} \epsilon^{-1} L^{\mathrm{Beta}}(\epsilon) \,/\, \epsilon = \frac{d}{d\epsilon} L^{\mathrm{Beta}}(\epsilon)\Big|_{\epsilon=0} = a(1-\epsilon)^{a-1}\Big|_{\epsilon=0} = a\,,$$

so $L_*^{\mathrm{Beta}} = a$. ∎

Theorem 9. *The total variation distance $\|\Gamma_{n+c} - \Gamma'_{n+c}\|$, between two copies of the Dirichlet process means Markov chain at time $n+c$, satisfies that*

$$\|\Gamma_{n+c} - \Gamma'_{n+c}\| \le (a \,/\, A)(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + (1-k_A)^c\,,$$

for any $A > 0$. Furthermore, if α_0 has density with respect to Lebesgue measure which is bounded by K, then

$$\|\Gamma_{n+c} - \Gamma'_{n+c}\| \le (a/A)(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + (2AK)^c\,,$$

and for any $z \ge 0$,

$$\|\Gamma_{n(1+z/\log n)} - \Gamma'_{n(1+z/\log n)}\| \le 2Kan(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + e^{-zn}\,.$$

Proof. Given that $|\Gamma_{m-1} - X_m| \ge A$, we have by Lemma 7 that the probability of failing to couple at the m^{th} step is

$$\le Q_{R_m} \le a|1 - R_m| = a\left|\frac{\Gamma_{m-1} - \Gamma'_{m-1}}{\Gamma'_{m-1} - X_m}\right| \le \frac{a}{A}|\Gamma_{m-1} - \Gamma'_{m-1}|\,.$$

Furthermore if $m \ge n$ and we haven't yet attempted to couple by time m, then $|\Gamma'_m - \Gamma_m| \le |\Gamma'_n - \Gamma_n|$, so if $|\Gamma_{m-1} - X_m| \ge A$, then the coupling probability is at least $k_A \frac{a}{A}|\Gamma'_n - \Gamma_n|$.

Hence, the probability that we fail to couple on all c attempts is at most the probability that we never have $|\Gamma_{m-1} - X_m| \ge A$ for any $n \le m \le n+c$, plus the probability that

$k_A L^{\mathrm{Beta}}(|\Gamma'_n - \Gamma_n| \,/\, A)$. (Actually it should be the *conditional* probability, conditional on having $|\Gamma_{q-1} - X_q| < A$ for $n \le q \le m-1$. But fortunately the *differences* $|\Gamma'_n - \Gamma_n|$ depend only on the Y_i, not the X_i, so this doesn't matter.)

That is, if T is the time when we finally successfully couple, then

$$\mathbf{P}(T \ge n+c) \le \mathbf{E}\left(L^{\mathrm{Beta}}(|\Gamma'_n - \Gamma_n|) \,/\, A\right) + (1-k_A)^c$$

$$\le (L^{\mathrm{Beta}}_* \,/\, A)\mathbf{E}\left(|\Gamma'_n - \Gamma_n|\right)) + (1-k_A)^c = (a \,/\, A)(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + (1-k_A)^c$$

(using that a product of n Beta$(a,1)$'s is Gamma(a,n), and the formula for the mean of a Gamma).

Furthermore, if α_0 has density with respect to Lebesgue measure which is bounded by K, then $k_A = \inf_{g \in \mathbf{R}} \mathbf{P}(|X-g| \ge A) \ge 1 - 2AK$. Then the bound becomes

$$\mathbf{P}(T \ge n+c) \le (a/A)(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + (2AK)^c \,. \tag{10}$$

One good choice is to set $A = 1/2Kn$, and $c = zn/\log n$. Then the bound (10) becomes

$$\mathbf{P}(T \ge n + (zn/\log n)) \le 2Kna(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + (1/n)^{zn/\log n}$$

$$= 2Kan(1 - \frac{1}{a+1})^n |\Gamma'_0 - \Gamma_0| + e^{-zn} \,. \qquad \blacksquare$$

Choosing z large enough, this theorem gives an asymptotic convergence rate as $n \to \infty$ of $(1 - \frac{1}{a+1})^n$.

Note that in (8), $\mathbf{E}[Y_n] = (1 - \frac{1}{a+1})$. Hence, the asymptotic rate in the theorem is essentially (aside from the non-linearity of the log function) the same rate that $|\Gamma'_n - \Gamma_n|$ goes pointwise to 0; see the discussion at the end of the Introduction.

Addendum re Dirichlet process means example.

For specific numerical comparison, suppose $\alpha_0 = \mathbf{Uniform}[0,1]$. Then with $\Gamma'_0 \sim \pi(\cdot)$, and with (say) $\Gamma_0 = 1/2$, we have that $\mathbf{E}|\Gamma_0 - \Gamma'_0| \le 1/2$. Also for $0 \le A \le 1/2$, we have $k_A = 1 - 2A$, giving that

$$\|\mathcal{L}(\Gamma_{k+c}) - \pi(\cdot)\| \;\le\; (a/2A)\left(\frac{a}{a+1}\right)^k + (2A)^c \,.$$

With $a = 1$, this is ≤ 0.01 if $A = 1/4$ and $k = 9$ and $c = 8$, giving a number of iterations of $k + c = 17$, comparable to the 13 iterations reported by Guglielmi and Tweedie (2000). On the other hand, with $a = 50$, our bound is ≤ 0.01 if $A = 1/4$ and $k = 467$ and $c = 11$, giving a number of iterations of $k + c = 478$, which is many orders of magnitude less than the 6.2×10^{14} iterations reported by Guglielmi and Tweedie (2000). With $a = 100$ the bounds of Guglielmi and Tweedie (2000) are even larger, but our bound is ≤ 0.01 if $A = 0.26$ and $k = 994$ and $c = 13$, giving a number of iterations of $k + c = 1007$, which is still quite reasonable.

Similarly, suppose α_0 is the standard normal distribution, with $a = 10$. Then with $\Gamma_0' \sim \pi(\cdot)$, and with (say) $\Gamma_0 = 0$, we have (since α_0 and π have the same mean) that $\mathbf{E}|\Gamma_0 - \Gamma_0'| = 1$. Also $k_A = 2\,\Phi(-A)$. Hence, we obtain that

$$\|\mathcal{L}(\Gamma_{k+c}) - \pi(\cdot)\| \;\leq\; (a/A)\left(\frac{a}{a+1}\right)^k + (1 - 2\,\Phi(-A))^c \;.$$

This is ≤ 0.01 if $A = 1$, $k = 75$, and $c = 17$, giving a number of iterations of $k + c = 92$, which is substantially less than the 10^6 iterations reported by Guglielmi and Tweedie (2000).

We thus see that, if a is at all large, then our one-shot coupling bound on the convergence rate of the Dirichlet process means Markov chain is significantly better than the standard coupling bound of Guglielmi and Tweedie (2000).

Remark. We could instead use the theory of large deviations, and write

$$|\Gamma_n - \Gamma_n'| = \exp(\sum_{i=1}^{n} \log Y_i) = \rho^n e^{nk(n)}$$

where $\rho = \exp \mathbf{E}[\log Y_i]$, and $k(n)$ is related to Large Deviations theory and satisfies that

$$\mathbf{P}(k(n) > \epsilon) \leq e^{-I(\epsilon)n}$$

(where I is the corresponding large deviations rate function), so that $\mathbf{P}\left(|\Gamma_n - \Gamma_n'| > \delta\right) \leq e^{-I(\epsilon)n}$ whenever $n > \frac{\log \delta}{\epsilon + \log \rho}$. For the Dirichlet means example considered in this section, we can compute $I(\epsilon)$ explicitly. Indeed, $I(\epsilon) = L^*(y)$ where $y = a^{-1} + \epsilon$, and $L^*(y) =$

$\sup_\lambda(\lambda y - \log M(\lambda))$ with $M(\lambda) = \mathbf{E}(e^{\lambda \log Y_i}) = \frac{a}{a-\lambda}$ [since the log of a Beta$(a,1)$ is Gamma$(a,1)$, i.e. Exp(a)]. The maximum occurs when $\lambda = a - \frac{1}{y}$, and we compute that $L^*(y) = ay - 1 - \log(ay)$, so that $I(\epsilon) = \epsilon a - \log(1+\epsilon a)$. It is possible to continue this analysis, however it gets messy and appears to yield a slower rate of convergence than the method presented herein.

Acknowledgements. We thank Richard Tweedie and Jeremy Quastel for helpful discussions, and thank the anonymous referee for corrections.

REFERENCES

L. Arnold and H. Crauel (1992), Iterated function systems and multiplicative ergodic theory. In *Diffusion Theorey and Related Problems in Analysis II*, M. Pinsky and V. Wihstatz, eds., 283–305. Birkhauser, Boston.

P. Billingsley (1995), Probability and Measure, 3rd ed. John Wiley & Sons, New York.

A.A. Borovkov and S.G. Foss (1992), Stochastically recursive sequences and their generalizations. Siberian. Adv. Math. **2**, 16–81.

P. Diaconis and D. Freedman (1999), Iterated random functions. SIAM Review **41**, 45–76.

L. Dubins and D. Freedman (1966), Invariant probabilities for certain Markov processes. Ann. Math. Stat. **37**, 837–844.

J. Elton (1990), A multiplicative ergodic theorem for Lipchitz maps. Stoch. Proc. Appl. **34**, 39–47.

P.D. Feigin and R.L. Tweedie (1989), Linear functionals and Markov chains associated with Dirichlet processes. Math. Proc. Cambridge Phil. Soc. **105**, 579–585.

G.B. Folland (1984), Real analysis: Modern techniques and their applications. John Wiley & Sons, New York.

S.G. Foss and R.L. Tweedie (1998). Perfect simulation and backward coupling. Stochastic Models **14**, 187–203.

A. Guglielmi and R.L. Tweedie (2000), MCMC estimation of the law of the mean of a Dirichlet process. Preprint, Dept. of Biostatistics, University of Minnesota.

S. Jarner and R.L. Tweedie (2000a), Locally contracting iterated random functions and stability of Markov chains. Preprint.

S. Jarner and R.L. Tweedie (2000b), Stability properties of Markov chains defined via iterated random functions. Preprint.

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

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

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

J.G. Propp and D.B. Wilson (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures and Algorithms **9**, 223–252.

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

G.O. Roberts and R.L. Tweedie (1999), Bounds on regeneration times and convergence rates for Markov chains. Stoch. Proc. Appl. **80**, 211–229.

G.O. Roberts and R.L. Tweedie (2000), Rates of convergence for stochastically monotone stochastic processes. J. Appl. Prob., to appear.

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