Elect. Comm. in Probab. **2** (1997) 13–25

ELECTRONIC COMMUNICATIONS in PROBABILITY

GEOMETRIC ERGODICITY AND HYBRID MARKOV CHAINS

GARETH O. ROBERTS
Statistical Laboratory, University of Cambridge
Cambridge CB2 1SB, U.K.
e-mail: G.O.Roberts@statslab.cam.ac.uk.

JEFFREY S. ROSENTHAL
Department of Statistics, University of Toronto
Toronto, Ontario
Canada M5S 3G3.
e-mail: jeff@utstat.toronto.edu.

first submitted:August 1, 1996; revised *April 11, 1997*

AMS 1991 Subject classification: 60J35 (60J10)
Keywords and phrases: Ergodicity, Markov Chain, Monte Carlo.

Abstract
Various notions of geometric ergodicity for Markov chains on general state spaces exist. In this paper, we review certain relations and implications among them. We then apply these results to a collection of chains commonly used in Markov chain Monte Carlo simulation algorithms, the so-called hybrid chains. We prove that under certain conditions, a hybrid chain will "inherit" the geometric ergodicity of its constituent parts.

1 Introduction

A question of increasing importance in the Markov chain Monte Carlo literature (Gelfand and Smith, 1990; Smith and Roberts, 1993) is the issue of geometric ergodicity of Markov chains (Tierney, 1994, Section 3.2; Meyn and Tweedie, 1993, Chapters 15 and 16; Roberts and Tweedie, 1996). However, there are a number of different notions of the phrase "geometrically ergodic", depending on perspective (total variation distance vs. in L^2; with reference to a particular V-function; etc.). One goal of this paper is to review and clarify the relationship between such differing notions.

We first discuss a general result (Proposition 1) giving the equivalence of a number of related ergodicity notions, involving total variation distance and V-uniform norms. Some of these equivalences follow from standard treatments of general state space Markov chains (Nummelin, 1984; Asmussen, 1987; Meyn and Tweedie, 1993), though they may not have previously been stated explicitly. Others of these equivalences are related to "solidarity properties" (Nummelin and Tweedie, 1978; Vere-Jones, 1962), regarding when a geometric rate of convergence can be chosen independently of the starting position.

We then turn to L^2 theory, and discuss (Theorem 2) a number of equivalences of geometric ergodicity of reversible Markov chains, on L^1 and L^2. The essence of our analysis is the spectral theorem (e.g. Rudin, 1991; Reed and Simon, 1972; Conway, 1985) for bounded self-adjoint operators on a Hilbert space. Again, we believe that these equivalences are known, though they may not have been explicitly stated in this way.

We further show that the conditions of Proposition 1 imply the conditions of Theorem 2. We are unable to establish the converse in complete generality, however we do note that the two sets of conditions are equivalent for most chains which arise in practice. We also argue (Corollary 3) that any of these equivalent conditions are sufficient to establish a functional central limit theorem for empirical averages of L^2 functions.

We then turn our attention (Section 3) to the geometric ergodicity of various "hybrid" Markov chains which have been suggested in the literature (Tierney, 1994, Section 2.4; Chan and Geyer, 1994; Green, 1994). After a few preliminary observations, we prove (Theorem 6) that under suitable conditions, hybrid chains will "inherit" the geometric ergodicity of their constituent chains. This suggests the possibility of establishing the geometric ergodicity of large and complicated Markov chain algorithms, simply by verifying the geometric ergodicity of the simpler chains which give rise to them.

We note that there are various alternatives to considering distributional convergence properties of Markov chains, such as considering the asymptotic variance of empirical estimators (cf. Geyer, 1992; Greenwood et al., 1995). But we do not pursue that here.

2 Equivalences of geometric ergodicity

We now turn our attention to results for geometric convergence. We consider a ϕ-irreducible, aperiodic Markov chain $P(x,\cdot)$ on a state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$. [Recall that a chain is *ϕ-irreducible* if there is a non-zero measure ϕ on $\mathcal{X}$, such that if $\phi(A) > 0$, then $P_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$.] We shall sometimes assume that the underlying σ-algebra on $\mathcal{X}$ is countably generated; this is assumed in much of general state space Markov chain theory (cf. Jain and Jamison, 1967; Meyn and Tweedie, 1993, pp. 107 and 516), and is also necessary to ensure that the function $x \mapsto \|P(x,\cdot) - \pi(\cdot)\|_{\mathrm{var}}$ is measurable (see Appendix).

We recall that P acts to the right on functions and to the left on measures, so that

$$\mu P(A) = \int P(x,A)\,\mu(dx), \qquad Pf(x) = \int f(y)\,P(x,dy).$$

Call a subset $S \subseteq \mathcal{X}$ *hyper-small* if $\pi(S) > 0$ and there is $\delta_S > 0$ and $k \in \mathbb{N}$ such that $\frac{d\left(P^k(x,\cdot)\right)}{d\pi} \geq \delta_S \mathbf{1}_S(x)$, where $\frac{d\left(P^k(x,\cdot)\right)}{d\pi}$ is the density with respect to π of the absolutely continuous component of $P^k(x,\cdot)$. A fundamental result of Jain and Jamison (1967), see also Orey (1971), states that on a countably generated measure space, every set of positive π-measure contains a hyper-small set.

Following Meyn and Tweedie (1993, p. 385), given a positive function $V : \mathcal{X} \to \mathbb{R}$, we let L_V^∞ be the vector space of all functions from $\mathcal{X}$ to $\mathbb{R}$ for which the norm $|f|_V = \sup\limits_{x\in\mathcal{X}} \frac{|f(x)|}{V(x)}$ is finite.

We let $L_{V,0}^\infty = \{f \in L_V^\infty \, ; \, \int_{\mathcal{X}} f d\pi = 0\}$. We denote operator norms as usual, viz.

$$\|T\|_{L_V^\infty} = \sup_{\substack{f \in L_V^\infty \\ |f|_V = 1}} |Tf|_V \, ; \qquad \|T\|_{L_{V,0}^\infty} = \sup_{\substack{f \in L_{V,0}^\infty \\ |f|_V = 1}} |Tf|_V \, .$$

Finally, we write $\mu(f)$ for $\int_{\mathcal{X}} f d\mu$, and let Π be the operator defined by $\Pi(f)(x) = \pi(f)$.
The following proposition borrows heavily from Meyn and Tweedie (1993) and Nummelin and Tweedie (1978). Note that the equivalence of (i) and (i') below shows that the geometric rate ρ may be chosen independently of the starting point $x \in \mathcal{X}$, though in general (i.e. for chains that are not "uniformly ergodic"), the multiplicative constants C_x will depend on x. Most of the remaining conditions concern the existence and properties of a geometric drift function V.

Proposition 2.1 *The following are equivalent, for a ϕ-irreducible, aperiodic Markov chain $P(x, \cdot)$ on a countably generated state space $\mathcal{X}$, with stationary distribution $\pi(\cdot)$.*

(i) The chain is π-a.e. geometrically ergodic, i.e. there is $\rho < 1$, and constants $C_x < \infty$ for each $x \in \mathcal{X}$, such that for π-a.e. $x \in \mathcal{X}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{var}} \;\leq\; C_x \, \rho^n \, , \qquad n \in I\!N,$$

where $\|\mu(\cdot)\|_{\mathrm{var}} = \frac{1}{2} \sup_{|f| \leq 1} |\mu(f)| = \sup_{A \subseteq \mathcal{X}} |\mu(A)|$;

(i') There are constants $\rho_x < 1$ and $C_x < \infty$ for each $x \in \mathcal{X}$, such that for π-a.e. $x \in \mathcal{X}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{var}} \;\leq\; C_x \, \rho_x^n \, , \qquad n \in I\!N;$$

(i") There is a hyper-small set $S \subseteq \mathcal{X}$, and constants $\rho_S < 1$ and $C_S < \infty$, such that

$$\|\int_S \frac{\pi(dy)}{\pi(S)} P^n(y, \cdot) - \pi(\cdot)\|_{\mathrm{var}} \;\leq\; C_S \, \rho_S^n \, , \qquad n \in I\!N;$$

(ii) There exists a π-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, which may be taken to have $\pi(V^j) < \infty$ for any fixed $j \in I\!N$, such that the chain is V-uniformly ergodic, i.e. for some $\rho < 1$ and some fixed constant $C < \infty$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_V \;\leq\; C \, V(x) \, \rho^n \, , \qquad x \in \mathcal{X} \, , \quad n \in I\!N,$$

where $\|\mu(\cdot)\|_V = \sup_{|f| \leq V} |\mu(f)|$;

(iii) There exists a π-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, which may be taken to have $\pi(V^j) < \infty$ for any fixed $j \in I\!N$, and a positive integer n, such that

$$\|P^n - \Pi\|_{L_V^\infty} \;<\; 1 \, ;$$

(iv) There exists a π-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, which may be taken to have $\pi(V^j) < \infty$ for any fixed $j \in I\!N$, and a positive integer n, such that

$$\|P^n\|_{L_{V,0}^\infty} \;<\; 1 \, .$$

Remark 2.1 *1. By the spectral radius formula (e.g. Rudin, 1991, Theorem 10.13), $r(T) = \inf_{n \geq 1} \|T^n\|^{1/n}$, so that the displayed equations in (iii) and (iv) above can be restated as $r(P - \Pi) < 1$ and $r(P\big|_{L^\infty_{V,0}}) < 1$, respectively.*

2. We may use the same function V for each of (ii), (iii), and (iv) above.

Proof:
$(i) \Longrightarrow (i')$: Immediate.
$(i') \Longrightarrow (i'')$: We note that ρ_x and C_x may not be measurable as functions of $x \in \mathcal{X}$. However, since $\mathcal{X}$ is countably generated, the function $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}}$ is measurable (see Appendix). Thus, we can define $r_x = \liminf_{n\to\infty} \|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}}^{1/n}$ and $K_x = \sup_n \|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}} r_x^{-n/2}$ as π-a.e.-finite measurable functions with $\pi(r_x < 1) = 1$. Then we can find $r < 1$ and $K < \infty$ so that the set $B = \{x \in \mathcal{X}\,;\, r_x \leq r,\, K_x \leq K\}$ has $\pi(B) > 0$. By Jain and Jamison (1967), there is a hyper-small subset $S \subseteq B$. For $x \in S$, we have $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}} \leq K r^n$, so therefore $\|\int_S \frac{\pi(dy)}{\pi(S)} P^n(y,\cdot) - \pi(\cdot)\|_{\text{var}} \leq K r^n$, as required.
$(i'') \Longrightarrow (i)$: This is the content of Theorem 1 of Nummelin and Tweedie (1978), which generalizes the countable state space results of Vere-Jones (1962).
$(i) \Longleftrightarrow (ii)$: Obviously (ii) implies (i). That (i) implies (ii) follows from Meyn and Tweedie (1993). Indeed, the existence of such a V follows from their Theorem 15.0.1 (i) and Theorem 5.2.2, the finiteness of $\mathbf{E}_\pi(V)$ follows from their Theorem 15.0.1 (iii) and Theorem 14.3.7 (see also Meyn and Tweedie, 1994, Proposition 4.3 (i)), and the possibility of having $\mathbf{E}_\pi(V^j) < \infty$ follows from their Lemma 15.2.9.
$(ii) \Longleftrightarrow (iii)$: Clearly, (ii) is equivalent to the existence of specified V, ρ, and C, such that

$$\sup_{|f| \leq V} \sup_{x \in \mathcal{X}} \frac{|(P^n f)(x) - \pi(f)|}{V(x)} \;\leq\; C\rho^n\,, \qquad n \in \mathbb{N}\,,$$

or

$$\|P^n - \Pi\|_{L^\infty_V} \;\leq\; C\rho^n\,, \qquad n \in \mathbb{N}\,.$$

Hence if $n > \frac{\log C}{\log(1/\rho)}$, then $\|P^n - \Pi\|_{L^\infty_V} < 1$, proving (iii). For the converse, if $\|P^n - \Pi\|_{L^\infty_V} = s < 1$, then by sub-multiplicity of operator norms, $\|P^{nk} - \Pi\|_{L^\infty_V} \leq s^k$, and (ii) follows.
$(iii) \Longleftrightarrow (iv)$: Clearly (iii) implies (iv). For the converse, if $\|P^n\|_{L^\infty_{V,0}} = \rho < 1$, and $|f|_V \leq 1$, then

$$|(P^{nk} - \Pi)f|_V \;=\; |P^{nk}(f - \pi(f))|_V \;\leq\; \rho^k |f - \pi(f)|_V \;\leq\; \rho^k(1 + \pi(V))\,.$$

Taking $j = 1$ so that $\pi(V) < \infty$, we see that this expression will be less than 1 for sufficiently large k, proving (iii). □

We now turn attention to results involving the spectral theory of bounded self-adjoint operators on Hilbert spaces.
Given a Markov operator P with stationary distribution π, a number $1 \leq p < \infty$, and a signed measure μ on $\mathcal{X}$, define $\|\mu\|_{L^p(\pi)}$ by

$$\|\mu\|_{L^p(\pi)}^p = \begin{cases} \int_{\mathcal{X}} \left|\frac{d\mu}{d\pi}\right|^p d\pi & \mu << \pi \\ \mu^+(\mathcal{X}) + \mu^-(\mathcal{X}), & p = 1 \\ \infty, & \text{otherwise} \end{cases}$$

(note that if $\mu << \pi$ and $p = 1$, the two definitions coincide), set $L^p(\pi) = \{\mu\, ;\, \|\mu\|_{L^p(\pi)} < \infty\}$, and set $\|P\|_{L^p(\pi)} = \sup\{\|\mu P\|_{L^p(\pi)}\, ;\, \|\mu\|_{L^p(\pi)} = 1\}$. It is well known (see e.g. Baxter and Rosenthal, 1995, Lemma 1) that we always have $\|P\|_{L^p(\pi)} \leq 1$.
Recall that P is *reversible* with respect to π if $\pi(dx)P(x, dy) = \pi(dy)P(y, dx)$ as measures on $\mathcal{X} \times \mathcal{X}$; this is equivalent to P being a self-adjoint operator on the Hilbert space $L^2(\pi)$, with inner product given by $\langle \mu, \nu\rangle = \int_{\mathcal{X}} \frac{d\mu}{d\pi}\frac{d\nu}{d\pi}\, d\pi$.
If P is reversible w.r.t. π, and $\frac{d\mu}{d\pi} = f$, then $\frac{d(\mu P)}{d\pi} = Pf$, as is easily checked by writing out the definitions and using reversibility. In particular, the action of P on signed measures $\mu \in L^2(\pi)$ is precisely equivalent to the action of P on functions f with $\pi(f^2) < \infty$. We shall apply the usual spectral theory (e.g. Rudin, 1991; Reed and Simon, 1972; Conway, 1985) to the operator P acting on measures in $L^2(\pi)$.

Theorem 2.1 *Let P be a Markov operator on a state space $\mathcal{X}$, reversible with respect to the probability measure π (so that P is self-adjoint on $L^2(\pi)$). Then the following are equivalent, and furthermore are all implied by any of the equivalent conditions of Proposition 2.1.*

(i) P is $L^2(\pi)$-geometrically ergodic, i.e. there is $\rho < 1$ such that for each probability measure $\mu \in L^2(\pi)$, there is $C_\mu < \infty$ with

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} \ \leq\ C_\mu\, \rho^n\,, \qquad n \in \mathbb{N};$$

(ii) P has an $L^2(\pi)$ spectral gap, i.e. there is $\rho < 1$ such that for each signed measure $\mu \in L^2(\pi)$ with $\mu(\mathcal{X}) = 0$,

$$\|\mu P(\cdot)\|_{L^2(\pi)} \ \leq\ \rho\, \|\mu\|_{L^2(\pi)}\,;$$

(iii) $P\big|_{L^2(\pi)}$ is geometrically ergodic, i.e. there is $\rho < 1$ such that for each probability measure $\mu \in L^2(\pi)$, there is $C_\mu < \infty$ such that

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{var}} \ \leq\ C_\mu\, \rho^n\,, \qquad n \in \mathbb{N};$$

(iv) $P\big|_{L^2(\pi)}$ is geometrically ergodic, with $C_\mu = \frac{1}{2}\|\mu - \pi\|_{L^2(\pi)}$, i.e. there is $\rho < 1$ such that for each probability measure $\mu \in L^2(\pi)$,

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{var}} \ \leq\ \frac{1}{2}\|\mu - \pi\|_{L^2(\pi)}\, \rho^n\,, \qquad n \in \mathbb{N};$$

Remark 2.2 *We may use the same value of ρ in each of these four equivalent conditions. Also note that, by self-adjointness, in condition (ii) above it is equivalent to say that $\|P\big|_{\pi^\perp}\|_{L^2(\pi)} \leq \rho < 1$ or that the spectral radius satisfies $r(P) \leq \rho < 1$ (cf. Conway, 1985, Proposition VIII.1.11(e)); we make use of this below. Finally, note that for probability measures μ, we have $\|\mu - \pi\|^2_{L^2(\pi)} = \|\mu\|^2_{L^2(\pi)} - 1$.*

Proof:
$(ii) \Longrightarrow (i)$, and $(iv) \Longrightarrow (iii)$: Immediate.
$(i) \Longrightarrow (iii)$, and $(ii) \Longrightarrow (iv)$: These follow from the Cauchy-Schwarz inequality, since $\|\mu\|_{\mathrm{var}} = \frac{1}{2}\|\mu\|_{L^1(\pi)} \leq \frac{1}{2}\|\mu\|_{L^2(\pi)}$.
$(iii) \Longrightarrow (ii)$: By the triangle inequality, condition (iii) implies that for any signed measure $\mu \in L^2(\pi)$, there is $C_\mu < \infty$ such that

$$\|\mu P^n(\cdot) - \mu(\mathcal{X})\, \pi(\cdot)\|_{\mathrm{var}} \ \leq\ C_\mu \rho^n\,, \qquad n \in \mathbb{N}\,. \qquad (*)$$

Let $\mathcal{E}(\cdot)$ be the spectral measure corresponding to P acting on $L^2(\pi)$ (see e.g. Rudin, 1991; Reed and Simon, 1972; Conway, 1985), so that $\mu P^k = \int_{-1}^{1} \lambda^k \mu\mathcal{E}(d\lambda)$ for all $k \in \mathbb{N}$ and $\mu \in L^2(\pi)$. Choose $r \in (\rho, 1)$.

If (ii) does not hold, then either $\mathcal{E}((-1,-r))$ or $\mathcal{E}((r,1))$ is not the zero operator. Assume (by replacing P by P^2 if necessary) that $\mathcal{E}((r,1)) \neq 0$. Hence there is a non-zero signed measure ν in the range of $\mathcal{E}((r,1))$. Since P is self-adjoint, $\nu(\mathcal{X}) = \langle \pi, \nu \rangle = 0$ by orthogonality of the spectral projections, hence there is $\mathcal{Y} \subseteq \mathcal{X}$ with $\nu(\mathcal{Y}) = \nu\mathcal{E}\left((r,1)\right)(\mathcal{Y}) > 0$. Define the signed measure m by $m(S) = \nu\mathcal{E}(S)(\mathcal{Y})$ for Borel $S \subseteq (r,1)$. Let $(r,1) = A^+ \cup A^-$ be the Hahn decomposition for m (see e.g. Royden, 1968, pp. 235–236). Since $m(r,1) = m(A^+) + m(A^-) = \nu(\mathcal{Y}) > 0$, and $m(A^-) \leq 0$, therefore $m(A^+) > 0$. Set $\omega = \nu\mathcal{E}(A^+) \in L^2(\pi)$, so that $\omega << \pi$ and $\omega(\mathcal{Y}) = m(A^+) > 0$.

Again by orthogonality, $\omega(\mathcal{X}) = 0$. On the other hand, using the functional calculus equation given in Conway (1985, p. 265, equation 1.11), and using that for $d\lambda \subseteq A^+$ we have $\omega\mathcal{E}(d\lambda) = \nu\mathcal{E}(d\lambda)$, we see that for $k \in \mathbb{N}$,

$$
\begin{aligned}
2\|\omega P^k(\cdot)\|_{\mathrm{var}} &= \|\omega P^k(\cdot)\|_{L^1(\pi)} \\
&= \int_{\mathcal{X}} \left| \frac{d(\omega P^k)}{d\pi}(x) \right| \pi(dx) \\
&\geq \int_{\mathcal{Y}} \left| \frac{d(\omega P^k)}{d\pi}(x) \right| \pi(dx) \\
&\geq \int_{\mathcal{Y}} \frac{d(\omega P^k)}{d\pi}(x) \pi(dx) \\
&= \langle\ \pi\big|_{\mathcal{Y}},\ \omega P^k\ \rangle \\
&= \langle\ \pi\big|_{\mathcal{Y}},\ \omega \int_{A^+} \lambda^k \mathcal{E}(d\lambda)\ \rangle \\
&= \int_{A^+} \lambda^k\ \langle\ \pi\big|_{\mathcal{Y}},\ \omega\mathcal{E}(d\lambda)\ \rangle \\
&= \int_{A^+} \lambda^k \Big[\int_{\mathcal{Y}} \frac{d(\omega\mathcal{E}(d\lambda))}{d\pi}(x)\ \pi(dx) \Big] \\
&= \int_{A^+} \lambda^k \left(\omega\mathcal{E}(d\lambda)\right)(\mathcal{Y}) \\
&= \int_{A^+} \lambda^k \left(\nu\mathcal{E}(d\lambda)\right)(\mathcal{Y}) \\
&= \int_{A^+} \lambda^k m(d\lambda) \\
&\geq r^k m(A^+)
\end{aligned}
$$

which contradicts $(*)$.

Relation to Proposition 2.1: It remains to show that these conditions are implied by the equivalent conditions of Proposition 2.1. We show that condition (iii) is implied by Proposition 2.1 part (ii) (with $j=2$). Indeed, for $\mu \in L^2(\pi)$, we have (using the triangle inequality and the Cauchy-Schwartz inequality) that

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{var}} \le \|\mu P^n(\cdot) - \pi(\cdot)\|_V \le \int_{\mathcal{X}} \|P^n(x,\cdot) - \pi(\cdot)\|_V \, \mu(dx)$$

$$\le \int_{\mathcal{X}} CV(x)\rho^n \mu(dx) = C\rho^n \int_{\mathcal{X}} V(x) \frac{d\mu}{d\pi}(x)\, \pi(dx) \le C\rho^n \sqrt{\mathbf{E}_\pi(V^2)\mathbf{E}_\pi\left(\left(\frac{d\mu}{d\pi}\right)^2\right)}\,.$$

Since $\mathbf{E}_\pi(V^2)\mathbf{E}_\pi\left(\left(\frac{d\mu}{d\pi}\right)^2\right) < \infty$, condition (iii) follows. □

Remark 2.3 *1. For most reversible chains which arise, the conditions of Proposition 1 and Theorem 2 are actually equivalent. For example, this is obviously true if the point masses δ_x are all in $L^2(\pi)$ (which holds for an irreducible chain on a discrete measure space), or if $P(x,\cdot) \in L^2(\pi)$ for all x. More generally, if the chain is of the form $P(x,\cdot) = (1-a_x)\delta_x + a_x \nu_x(\cdot)$ where $\nu_x \in L^2(\pi)$ and $a_x > 0$, then the two sets of conditions are again seen to be equivalent, since if condition (iii) of Theorem 2 holds, then condition (i') of Proposition 1 also holds. This includes most examples of Metropolis-Hastings algorithms (Metropolis et al., 1953; Hastings, 1970). However, we do not know if these conditions are equivalent in complete generality.*
2. We note that a number of other mixing conditions for Markov chains are available, though we do not pursue them here. See for example Rosenblatt (1962, Sections V b and VIII d), and Carmona and Klein (1983).

Finally, we consider the existence of central limit theorems for empirical averages $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$ of a function $g : \mathcal{X} \to \mathbb{R}$ evaluated at the realized values $X_1, X_2, \ldots$ of a Markov chain run. Such limit theorems are often considered in Markov chain Monte Carlo applications (cf. Tierney, 1994, Theorem 4). Geyer (1992, Section 2) and Chan and Geyer (1994) point out the usefulness of a certain Markov Functional Central Limit Theorem result (see e.g. Bhattacharya, 1982, Theorem 2.1; Kipnis and Varadhan, 1986, Corollary 1.5), which states the following. For a Markov chain $P(\cdot,\cdot)$ on $\mathcal{X}$, reversible with respect to $\pi(\cdot)$, if $\pi(g) = 0$ and the quantity

$$\sigma_g^2 \;=\; \int_{-1}^{1} \frac{1+\lambda}{1-\lambda}\, e_g(d\lambda)$$

is finite, where $e_g(S) = \int_{\mathcal{X}} g(x)\left(\mathcal{E}(S)g\right)(x)\pi(dx)$, then the continuous-time processes $Y_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor} g(X_j)$ converge weakly to Brownian motion with variance σ_g^2.
On the other hand, if the Markov operator P has spectral radius $\rho < 1$, as in condition (ii) of Theorem 2, then it is easily seen (cf. Geyer, 1992, Section 3.5) that for $\pi(g^2) < \infty$,

$$\sigma_g^2 \;\le\; \frac{1+\rho}{1-\rho}\,\pi(g^2) \;<\; \infty\,,$$

so that the above Markov Functional Central Limit Theorem applies. We thus obtain:

Corollary 2.1 *Let the reversible Markov operator P satisfy any one equivalent condition of Proposition 1 or of Theorem 2 (e.g., let P be geometrically ergodic). Let $g : \mathcal{X} \to I\!R$ with $\pi(g) = 0$ and $\pi(g^2) < \infty$. Then $\sigma_g^2 < \infty$, and as $n \to \infty$, the processes $Y_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} g(X_j)$ converge weakly (in the Skorohod topology, on any finite interval) to Brownian motion with variance σ_g^2. In particular, setting $t = 1$, the random variable $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i)$ converges in distribution to the normal distribution $N(0, \sigma_g^2)$.*

This corollary directly generalizes Chan and Geyer (1994, Theorem 2) for reversible chains; in particular, it shows that their condition $\pi\left(|g|^{2+\epsilon}\right) < \infty$ is then unnecessary. See Chan and Geyer (1994), and references therein, for some other approaches to central limit theorems for Markov chains, including those involving drift conditions (cf. Meyn and Tweedie, 1993, Theorem 17.5.4).

Remark 2.4 *We note that the Markov Functional Central Limit Theorem used above applies only to a reversible Markov process with a spectral gap. For more general reversible Markov processes, the less restrictive results of Kipnis and Varadhan (1986) may be very useful. Furthermore, we note that corresponding results for non-reversible processes are an active area of study, for example in the contexts of interacting particle systems and in homogenization theory; see for example Derriennic and Lin (1996), and the recent unpublished works of Papanicolaou, Komorovskii, Carmona, Xu, Lamdin, Olla, Yau, and Sznitman.*

3 Geometric ergodicity of hybrid samplers

Given a probability distribution $\pi(\cdot)$ on the state space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_k$, the usual deterministic-scan Gibbs sampler (DUGS) is the Markov kernel $P = P_1 P_2 \ldots P_k$, where P_i is the Markov kernel which replaces the i^{th} coordinate by a draw from $\pi(dx_i | x_{-i})$, leaving x_{-i} fixed (where $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$). This is a standard Markov chain Monte Carlo technique (Gelfand and Smith, 1990; Smith and Roberts, 1993; Tierney, 1994). The random-scan Gibbs sampler (RSGS), given by $P = \frac{1}{k} \sum_i P_i$, is sometimes used instead.
Often the full conditionals $\pi(dx_i | x_{-i})$ may be easily sampled, so that DUGS may be efficiently run on a computer. However, sometimes this is not feasible. Instead, one can define new operators Q_i which are easily sampled , such that Q_i^n converges to P_i as $n \to \infty$. This is the method of "one-variable-at-a-time Metropolis-Hastings" or "Metropolis within Gibbs" (cf. Tierney, 1994, Section 2.4; Chan and Geyer, 1994, Theorem 1; Green, 1994).
Such samplers prompt the following definition.

Definition 3.1 *Let $C = (Q_1, Q_2, \ldots, Q_k)$ be a collection of Markov kernels on a state space $\mathcal{X}$. The random-scan hybrid sampler for C is the sampler defined by*

$$P_C = \frac{1}{k}(Q_1 + \ldots + Q_k).$$

A common example is the variable-at-a-time Metropolis-Hastings algorithms mentioned above. For another example, if the Q_i are themselves Gibbs samplers, then the random-scan hybrid sampler would correspond to building a large Gibbs sampler out of smaller ones. Similarly, if the Q_i are themselves Metropolis-Hastings algorithms (perhaps with singular proposals, cf.

Tierney, 1995), then the random-scan hybrid sampler is also a Metropolis-Hastings algorithm but with a different proposal distribution. It is important to note that this usage of the word "hybrid" is distinct from the more specific notion of "hybrid Monte Carlo" algorithms as studied in the physics literature (cf. Duane et al., 1987; Neal, 1993, Section 5.2).
We now discuss some observations about geometric convergence of these samplers. Our first result is a cautionary example.

Proposition 3.1 *Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, and let π be a probability measure on $\mathcal{X}$. Let Q_1 and Q_2 be two Markov operators on $\mathcal{X}$ which fix the second and first coordinates (respectively). Assume that the usual RSGS for π is geometrically ergodic. Assume further that for each fixed $y \in \mathcal{X}_2$ [resp. $x \in \mathcal{X}_1$], we have that Q_1 [resp. Q_2] restricted to $\mathcal{X}_1 \times \{y\}$ [resp. $\{x\} \times \mathcal{X}_2$] is geometrically ergodic, with stationary distribution equal to $\pi(dx|y)$ [resp. $\pi(dy|x)$]. Then it is still possible that the hybrid sampler $P_C = \frac{1}{2}(Q_1 + Q_2)$ fails to be geometrically ergodic on $\mathcal{X}$.*

Proof:
A result of Roberts and Tweedie (1996) states that a Metropolis algorithm is not geometrically ergodic if the infimum of acceptance probabilities is 0. Therefore consider the following example. Let π be the bivariate density of two independent standard normal components. By independence, RSGS is geometrically ergodic (in fact uniformly ergodic with rate 1/2). Now let Q_1 be the following random walk Metropolis algorithm. Given $(X_n, Y_n) = (x, y)$, let $Y_{n+1} = y$ and propose a candidate Z_{n+1} for X_{n+1} from a $N(x, 1 + y^2)$ distribution, accepting with the probability: $1 \wedge \pi(Z_{n+1})/\pi(X_n)$. Similarly define Q_2 on fixed x by proposing a candidate for Y_{n+1} according to $N(y, 1 + x^2)$. For fixed y, Q_1 is geometrically ergodic; and for fixed x, Q_2 is geometrically ergodic (see for example Roberts and Tweedie, 1996). However it is easy to verify that along any sequence of points (x_i, y_i) such that $|x_i| \to \infty$ and $|y_i| \to \infty$, the acceptance probability for each of Q_1 and Q_2 (and hence also for P_C) goes to 0. This provides our required counterexample. □
To continue, we need some notation. Let π be a probability measure on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_k$, and let P_i be the operator which replaces the i^{th} coordinate of $x \in \mathcal{X}$ by a draw from $\pi(dx_i \,|\, x_{-i})$, leaving x_{-i} unchanged. Let $\mathcal{M}(\mathcal{X})$ be the space of all σ-finite signed measures on $\mathcal{X}$, and let $\mathcal{M}_0 = \{\mu \in \mathcal{M}(\mathcal{X})\,;\, \mu(\mathcal{X}) = 0\}$. Given a norm $\mathcal{N}$ on $\mathcal{M}(\mathcal{X})$, and a Markov operator P on $\mathcal{X}$, set $\|P\|_{\mathcal{N}} = \sup\limits_{\mathcal{N}(\mu)=1} \mathcal{N}(\mu P)$. Say P is *$\mathcal{N}$-geometrically ergodic* if $\pi P = \pi$ and there is a positive integer m with $\|P^m|_{\mathcal{M}_0}\|_{\mathcal{N}} < 1$. (Note that, if $\mathcal{N}$ is total variation distance, then $\mathcal{N}$-geometric ergodicity is equivalent to uniform ergodicity. On the other hand, if $\mathcal{N}(\mu) = \|\mu\|_{L^2(\pi)}$, then $\mathcal{N}$-geometric ergodicity is equivalent to condition (ii) of Theorem 2 above.) Finally, call $\mathcal{N}$ *π-contracting* if it has the property that $\|P\|_{\mathcal{N}} \leq 1$ for any Markov operator P satisfying $\pi P = \pi$. Thus, the $L^p(\pi)$ norms are all π-contracting (see e.g. Baxter and Rosenthal, 1995, Lemma 1), though the norm L^∞_V is usually not.

Proposition 3.2 *If DUGS is $\mathcal{N}$-geometrically ergodic, where $\mathcal{N}$ is π-contracting, then RSGS is also $\mathcal{N}$-geometrically ergodic.*

Proof:
Suppose that $\|(P_1 \ldots P_k)^n|_{\mathcal{M}_0}\|_{\mathcal{N}} < 1$. Then, since

$$\left(\frac{1}{k}(P_1 + \ldots + P_k)\right)^{nk} = \frac{1}{k^{nk}} \sum_{i_1, \ldots, i_{nk}} P_{i_1} \ldots P_{i_{nk}}$$

includes a term $(P_1 \ldots P_k)^n$, it has $\mathcal{N}$-norm < 1 by the triangle inequality. The result follows. $\square$

We note that, in general, the converse to this last result is false. For example, even if RSGS is geometrically ergodic, DUGS may be periodic.

We can now prove

Theorem 3.1 *Suppose DUGS or RSGS is $\mathcal{N}$-geometrically ergodic, where $\mathcal{N}$ is π-contracting. Suppose further that there are operators Q_i such that*

$$\lim_{n\to\infty} \|Q_i^n - P_i\|_{\mathcal{N}} \;=\; 0\,,$$

i.e. Q_i^n converges to P_i in $\mathcal{N}$-norm as $n \to \infty$. Then the random-scan hybrid sampler $P_C = \frac{1}{k}(Q_1 + \ldots + Q_k)$ is $\mathcal{N}$-geometrically ergodic.

Remark 3.1 *1. If furthermore $\mathcal{N}$ is such that $\|\nu T\|_{\text{var}} \le C_\nu \|T\|_{\mathcal{N}}$, then we can conclude geometric convergence in the total variation norm (cf. Baxter and Rosenthal, 1995, Theorem 1 (b)).*
2. Note that the condition $\lim\limits_{n\to\infty} \|Q_i^n - P_i\|_{\mathcal{N}} = 0$ implies a certain uniformity of convergence, uniform over choices of x_{-i}.

Proof:
If DUGS is $\mathcal{N}$-geometrically ergodic, there is a positive integer m with

$$\|(P_1 \ldots P_k)^m\big|_{\mathcal{M}_0}\|_{\mathcal{N}} \;<\; 1\,.$$

Also, by assumption,

$$\|Q_i^n - P_i\|_{\mathcal{N}} \to 0, \qquad n \to \infty\,,$$

so

$$\|(Q_1^n Q_2^n \ldots Q_k^n)^m - (P_1 P_2 \ldots P_k)^m\big|_{\mathcal{M}_0}\|_{\mathcal{N}} \to 0\,, \qquad n \to \infty\,.$$

Hence, there is a positive integer N such that

$$\|(Q_1^N Q_2^N \ldots Q_k^N)^m\big|_{\mathcal{M}_0}\|_{\mathcal{N}} \;= \rho \;<\; 1\,.$$

But then expanding out $(P_C)^{Nm}$, and recalling that $\|Q_*\|_{\mathcal{N}} \le 1$ where Q_* is any product of Q_i's, we see that

$$\|(P_C)^{Nm}\big|_{\mathcal{M}_0}\|_{\mathcal{N}} \;\le\; \left(\frac{1}{k}\right)^{Nm} \left(k^{Nm} - 1 + \rho\right) \;<\; 1\,.$$

The $\mathcal{N}$-geometric ergodicity follows.
The result for RSGS follows similarly, from noting that

$$\begin{aligned}\left(\frac{1}{k}\sum Q_i\right)^{Nm} &= \frac{1}{k^{Nm}}\left(\sum_{i_1,\ldots,i_m} Q_{i_1}^N \ldots Q_{i_m}^N\right) \;+\; \text{leftover} \\ &= \frac{1}{k^N}\left(\frac{1}{n}\sum Q_i^N\right)^m \;+\; \text{leftover}\end{aligned}$$

where "leftover" has $\mathcal{N}$-norm $\le 1 - k^{-N}$. Hence $\|(P_C)^m\big|_{\mathcal{M}_0}\|_{\mathcal{N}} \le 1 - (1-\rho)k^{-N} < 1.$ $\square$

In particular, this result shows that if a hybrid sampler is constructed out of chains which are $L^2(\pi)$-geometrically ergodic, and if RSGS is $L^2(\pi)$-geometrically ergodic, then the resulting hybrid sampler will also be $L^2(\pi)$-geometrically ergodic. It may be possible to use this result iteratively, to build up larger and larger geometric chains by combining smaller ones.

4 Appendix: Measurability of $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}}$

In this appendix we prove that, on a countably generated state space, the quantity $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}}$ is measurable as a function of x. This fact was needed in the proof of Proposition 1 above.

Lemma 4.1 *Let $\nu(\cdot)$ be a finite signed measure on $(\mathcal{X}, \mathcal{F})$. Suppose that $\mathcal{F}_0$ is a set algebra generating $\mathcal{F}$. Then for all $S \in \mathcal{F}$, and for all $\epsilon > 0$, there is $S_0 \in \mathcal{F}_0$ with*

$$|\nu(S) - \nu(S_0)| < \epsilon\,.$$

Proof:
Take ν to be a measure (otherwise consider its positive and negative parts separately and use the triangle inequality). It follows from Doob (1994, Chapter IV, Section 3) that we can find $S_0 \in \mathcal{F}_0$ with $\nu(S \Delta S_0)$ arbitrarily small. The result follows. □

Proposition 4.1 *Let $\nu(x,\cdot)$ be a bounded signed measure on $(\mathcal{X}, \mathcal{F})$, where $\mathcal{F}$ is countably generated by the sequence of sets $\{A_1, A_2, \ldots\}$. Assume that $\nu(\cdot, A)$ is a measurable function for all sets A. Then $\sup\limits_{A \in \mathcal{F}} \nu(x, A)$ is measurable.*

Proof:
The proof proceeds by demonstrating that the supremum can be written as a supremum of a countable collection of measurable functions.
Therefore let $\mathcal{F}_n = \sigma\{A_1, A_2, \ldots, A_n\}$. Now fix $x \in \mathcal{X}$. Let A^* be the set that achieves $\sup\limits_{A} \nu(x, A)$ (possible by the Hahn decomposition, e.g. Royden, 1968). Notice that $\bigcup\limits_n \mathcal{F}_n$ is a set algebra generating $\mathcal{F}$. So for arbitrary $\epsilon > 0$, we can apply the previous lemma to find a set $A \in \bigcup\limits_n \mathcal{F}_n$ with $\nu(A^*) \geq \nu(A) \geq \nu(A^*) - \epsilon$. Hence,

$$\sup_{A \in \mathcal{F}} \nu(x, A) = \sup_n \sup_{A \in \mathcal{F}_n} \nu(x, A)\,.$$

But $\sup\limits_{A \in \mathcal{F}_n} \nu(x, A)$ is clearly a measurable function of $x \in \mathcal{X}$. The result follows. □
Setting $\nu(x, A) = P^n(x, A) - \pi(A)$, we see that $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{var}}$ is measurable, as claimed.

Acknowledgements:
We thank Charlie Geyer for a number of very useful comments regarding spectral theory and central limit theorems. We thank Alison Gibbs, Phil Reiss, Peter Rosenthal, and Richard Tweedie for very helpful discussions. We thank the referee and the editor for many excellent suggestions.

References

[1] S. Asmussen (1987), Applied probability and queues. John Wiley & Sons, New York.

[2] J.R. Baxter and J.S. Rosenthal (1995), Rates of convergence for everywhere-positive Markov chains. *Stat. Prob. Lett.* **22**, 333-338.

[3] R.N. Bhattacharya (1982), On the functional central limit theorem and the law of the iterated logarithm for Markov processes. Z. Wahrscheinlichkeitstheorie Verw. Gebiete **60**, 185-201.

[4] R. Carmona and A. Klein (1983), Exponential moments for hitting times of uniformly ergodic Markov processes. *Ann. Prob.* **11**, 648-655.

[5] K.S. Chan and C.J. Geyer (1994), Discussion paper. *Ann. Stat.* **22**, 1747-1758.

[6] J.B. Conway (1985), A course in functional analysis. Springer-Verlag, New York.

[7] Y. Derriennic and M. Lin (1996), Sur le théorème limite central de Kipnis et Varadhan pour les chaînes réversibles ou normales. *Comptes Rendus de l'Académie des Sciences, Série I,* **323**, No. 9, 1053-1058.

[8] J.L. Doob (1994), Measure theory. Springer-Verlag, New York.

[9] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth (1987), Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216-222.

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

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

[12] P.J. Green (1994), Reversible jump MCMC computation and Bayesian model determination. Tech. Rep., Dept. Mathematics, U. of Bristol.

[13] P.E. Greenwood, I.W. McKeague, and W. Wefelmeyer (1995), Outperforming the Gibbs sampler empirical estimator for nearest neighbour random fields. Tech. Rep., Dept. Mathematics, U. of British Columbia.

[14] W.K. Hastings (1970), Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

[15] N. Jain and B. Jamison (1967), Contributions to Doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.* **8**, 19–40.

[16] C. Kipnis and S.R.S. Varadhan (1986), Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104**, 1-19.

[17] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1091.

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

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

[20] R.M. Neal (1993), Probabilistic inference using Markov chain Monte Carlo methods. Tech. Rep. CRG-TR-93-1, Dept. Comp. Sci., U. of Toronto.

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

[22] E. Nummelin and R.L. Tweedie (1978), Geometric ergodicity and R-positivity for general Markov chains. *Ann. Prob.* **6**, 404-420.

[23] S. Orey (1971), Lecture notes on limit theorems for Markov chain transition probabilities. Van Nostrand Reinhold, London.

[24] M. Reed and B. Simon (1972), Methods of modern mathematical physics. Volume I: Functional analysis. Academic Press, New York.

[25] G.O. Roberts and R.L. Tweedie (1996), Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms.*Biometrika* **83**, 95-110.

[26] M. Rosenblatt (1962), Random Processes. Oxford University Press, New York.

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

[28] R.L. Royden (1968), Real analysis, 2^{nd} ed. MacMillan, New York.

[29] W. Rudin (1991), Functional Analysis, 2^{nd} ed. McGraw-Hill, New York.

[30] 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.

[31] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). *Ann. Stat.* **22**, 1701-1762.

[32] L. Tierney (1995), A note on Metropolis Hastings kernels for general state spaces. Tech. Rep. 606, School of Statistics, U. of Minnesota.

[33] R.L. Tweedie (1978), Geometric ergodicity and R-positivity for general Markov chains. *Ann. Prob.* **6**, 404–420.

[34] D. Vere-Jones (1962), Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford (2)* **13**, 7–28.