# Variance Bounding Markov Chains

by

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

(September 2006; revised April 2007.)

**Abstract.** We introduce a new property of Markov chains, called *variance bounding*. We prove that, for reversible chains at least, variance bounding is weaker than, but closely related to, geometric ergodicity. Furthermore, variance bounding is equivalent to the existence of usual central limit theorems for all $L^2$ functionals. Also, variance bounding (unlike geometric ergodicity) is preserved under the Peskun order. We close with some applications to Metropolis-Hastings algorithms.

**Key words and phrases.** Markov chain Monte Carlo, Metropolis-Hastings algorithm, central limit theorem, variance, Peskun order, geometric ergodicity, spectrum.

## 1. Introduction.

Markov chain Monte Carlo (MCMC) algorithms are widely used in statistics, physics, and computer science. Measures of how good an MCMC algorithm is include quantitative bounds on convergence to stationarity (e.g. [34], [35], [14], [15]), qualitative convergence rates such as geometric ergodicity (e.g. [40], [39], [32], [29]), the existence of central limit theorems (e.g. [7], [40], [3], [10], [13], [2]), and bounds on asymptotic variance of estimators (e.g. [41], [7], [22]).

In this paper, we introduce a new notion, *variance bounding*. Roughly, a Markov chain is variance bounding if the asymptotic variances for functionals with unit stationary variance are uniformly bounded (precise definitions are given below). We shall show that, for reversible chains at least, variance bounding is implied by geometric ergodicity, and conversely if $P$ is variance bounding then $aI + (1-a)P$ is geometrically ergodic for all $0 < a < 1$. More importantly, we shall prove that a reversible Markov chain is variance bounding if and only if all $L^2$ functionals satisfy a usual central limit theorem, indicating that variance bounding

[*]Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, England. Email: `g.o.roberts@lancaster.ac.uk`.

[**]Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: `jeff@math.toronto.edu`. Web: `http://probability.ca/jeff/` Supported in part by NSERC of Canada.

is in some sense the "right" definition to use. We also prove that variance bounding is preserved under the Peskun partial ordering ([26], [40]) on Markov chains. Finally, applications to Metropolis-Hastings algorithms are presented.

## 2. Variance Bounding.

Given a Markov chain kernel $P$ on a state space $(\mathcal{X}, \mathcal{F})$ with unique stationary distribution $\pi(\cdot)$, we let $\{X_n\}$ follow the kernel $P$ in stationarity, so that $\mathbf{P}[X_n \in A] = \pi(A)$ for all $A \in \mathcal{F}$ and $n \in \mathbf{N} \cup \{0\}$, and also $\mathbf{P}[X_n \in A \,|\, X_0, \ldots, X_{n-1}] = P(X_{n-1}, A)$ for all $A \in \mathcal{F}$ and all $n \in \mathbf{N}$. For a functional $h : \mathcal{X} \to \mathbf{R}$ (assumed throughout to be measurable), the stationary variance is given by $\mathbf{Var}_\pi(h) = \mathbf{E}[(h(X_0) - \mathbf{E}[h(X_0)])^2]$, and the asymptotic variance is given by

$$\mathbf{Var}(h, P) \;=\; \lim_{n\to\infty} \frac{1}{n} \mathbf{Var}\Big( \sum_{i=1}^{n} h(X_i) \Big) . \tag{1}$$

If the Markov chain $P$ is to be used to estimate the stationary expected value of $h$ by $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$, then $\mathbf{Var}(h, P)$ is a measure of the Monte Carlo uncertainty of the estimate. Thus, for MCMC algorithms, it is desirable to make $\mathbf{Var}(h, P)$ as small as possible (cf. [40], [7], [41], [23], [22]). This prompts the following definition.

**Definition.** $P$ is *variance bounding* if there is $K < \infty$ such that $\mathbf{Var}(h, P) \leq K\ \mathbf{Var}_\pi(h)$ for all $h : \mathcal{X} \to \mathbf{R}$. Equivalently, $P$ is *variance bounding* if $\sup\{\mathbf{Var}(h, P);\ h : \mathcal{X} \to \mathbf{R},\ \mathbf{Var}_\pi(h) = 1\} < \infty$.

Note that in the case where $\mathbf{Var}_\pi(h) = \infty$, the required inequality holds automatically for all $K$.

Variance bounding is a natural property, in that it offers some control over the asymptotic variances $\mathbf{Var}(h, P)$. We study its relation to more traditional MCMC properties below. For most of our results, we assume that $P$ is *reversible* with respect to $\pi(\cdot)$, i.e. that

$$\int_{x\in A} \pi(dx)\, P(x, B) \;=\; \int_{x \in B} \pi(dx)\, P(x, A)\,, \qquad A, B \in \mathcal{F}\,. \tag{2}$$

It follows from [16] (see also [3]) that for reversible chains and $L^2$ functionals, the limit in equation (1) always exists, though it may be infinite.

## 3. Relation to geometric ergodicity.

Recall that a Markov chain kernel $P$ with stationary distribution $\pi(\cdot)$ is *geometrically ergodic* if there is $\rho < 1$ and $M : \mathcal{X} \to [0, \infty]$ $\pi$-a.e. finite (i.e., such that $\pi\{x \in \mathcal{X} : M(x) < \infty\} = 1$), such that $|P^n(x, A) - \pi(A)| \leq M(x)\,\rho^n$ for all $A \in \mathcal{F}$, $n \in \mathbf{N}$, and $x \in \mathcal{X}$. Geometric ergodicity is an often studied property (e.g. [40], [39], [32], [29]), which leads to many useful results such as central limit theorems (see next section).

However, geometric ergodicity is an overly strong notion in that it requires, among other things, that the Markov chain be *aperiodic*. Since estimates of functionals, and their variances $\mathbf{Var}(h, P)$, are essentially unaffected by periodicity considerations, it seems inappropriate to demand aperiodicity. And indeed, many Markov chains are variance bounding despite being periodic (e.g. the Markov chain $P_1$ in Example 9 below).

We now explore the relation between geometric ergodicity and variance bounding. We first show that for reversible chains, variance bounding is strictly *weaker* than geometric ergodicity. (Proofs of all theorems are deferred until Section 7.)

**Theorem 1.** *If $P$ is reversible and geometrically ergodic, then $P$ is variance bounding.*

Next, we show that $P$ is variance bounding if and only if any mixture of $P$ with the identity is geometrically ergodic. We write $I$ for the identity kernel, i.e. the Markov chain which never moves, so that $I(x, \{x\}) = 1$ for all $x \in \mathcal{X}$.

**Theorem 2.** *If $P$ is reversible, then the following are equivalent:*
*(i) $P$ is variance bounding.*
*(ii) $aI + (1-a)P$ is geometrically ergodic for all $0 < a < 1$.*
*(iii) $aI + (1-a)P$ is geometrically ergodic for some $0 \leq a < 1$.*

**Corollary 3.** *If $P$ is reversible, then for any fixed $0 \leq a < 1$, the following are equivalent:*
*(i) $P$ is variance bounding.*
*(ii) $aI + (1-a)P$ is variance bounding.*

Section 6 below contains some applications of Theorems 1 and 2. We next note that if $P$ has holding probabilities uniformly bounded away from 0, then variance bounding and geometrically ergodic are equivalent:

**Theorem 4.** *If $P$ is reversible and $\inf_{x\in\mathcal{X}} P(x, \{x\}) > 0$, then $P$ is variance bounding if and only if $P$ is geometrically ergodic.*

As an application of Theorem 4, suppose $P$ represents a *random-walk Metropolis* or *systematic-scan Metropolis-within-Gibbs* algorithm on $\mathbf{R}^d$, with proposal increment densities positive in a neighbourhood of 0, whose target density $t$ is $C^1$ with $\|\nabla \log t(x)\| \geq \delta > 0$ for all $x \in \mathcal{X}$. It then follows as in [33] that the rejection probabilities $P(x, \{x\})$ are uniformly bounded away from 0. Hence, by Theorem 4, variance bounding is equivalent to geometric ergodicity in this case.

Similarly, the two notions are equivalent if the operator $P$ is *positive*, i.e. if $\mathbf{E}[f(X_0)\, f(X_1)] \geq 0$ for all measurable $f : \mathcal{X} \to \mathbf{R}$ when $\{X_n\}$ is in stationarity:

**Theorem 5.** *If $P$ is reversible and positive, then $P$ is variance bounding if and only if $P$ is geometrically ergodic.*

As an application of Theorem 5, suppose $P$ represents a *data augmentation* algorithm, i.e. the $x$-coordinate (only) of a two-variable Gibbs sampler. It follows from Lemmas 3-1 and 3-2 of [18] that $P$ is reversible and positive. Hence, by Theorem 5, variance bounding is equivalent to geometric ergodicity in this case as well. (See also [11].)

In particular, the *slice sampler* (e.g. [25], [30], [24]) can be viewed as the $x$-coordinate of a two-variable Gibbs sampler. (This holds for product slice samplers as well, since the multiple auxiliary variables are conditionally independent and can be regarded as a single auxiliary vector.) So, for any slice sampler, variance bounding is equivalent to geometric ergodicity. For example, it is known [30] that the slice sampler is geometrically ergodic whenever $Q'(y)\, y^{1+\frac{1}{\alpha}}$ is non-increasing near 0, for some $\alpha > 1$, where $Q(y)$ is the measure of the set where the target density value is at least $y$. It follows immediately that the slice sampler is also variance bounding under these conditions.

In general, if $P$ is variance bounding, then a slight modification of $P$ is geometrically ergodic. Specifically, following [36], let $\overline{P}^n$ be the *binomial modification* of $P$, corresponding to doing an (independently chosen) random number $B_n$ of steps from $P$, where $B_n \sim \text{Binomial}(2n, 1/2)$. Thus, $\overline{P}^n = 2^{-2n} \sum_{i=0}^{2n} \binom{2n}{i} P^i$. Call $\overline{P}$ geometrically ergodic if, as usual, there is $\rho < 1$ and $\pi$-a.e. finite $M : \mathcal{X} \to [0, \infty]$ such that $|\overline{P}^n(x, A) - \pi(A)| \leq M(x)\, \rho^n$ for all $A \in \mathcal{F}$, $n \in \mathbf{N}$, and $x \in \mathcal{X}$. Then we have the following.

**Theorem 6.** *If $P$ is reversible, then $P$ is variance bounding if and only if $\overline{P}$ is geometrically ergodic.*

**Remark.** The stationary processes literature (e.g. [12], [2], [13]) defines many other mixing conditions such as $\alpha$-mixing, $\beta$-mixing, $\rho$-mixing, $\phi$-mixing, etc. These conditions are related to usual Markov chain ergodicity conditions, e.g. $\phi$-mixing is equivalent to uniform ergodicity, exponentially-fast $\beta$-mixing is equivalent to geometric ergodicity, $\alpha$-mixing is implied by Harris ergodicity, etc. However, none of these mixing conditions is implied by variance bounding, since the mixing conditions all require ergodicity, whereas periodic (and therefore non-ergodic) chains can still be variance bounding.

# 4. Relation to Central Limit Theorems.

An important issue in MCMC is the existence of central limit theorems (e.g. [7], [40], [3], [10], [13], [2]). Where central limit theorems are known to hold, they underpin practical MCMC strategies for Monte Carlo error assessment (see e.g. [8]).

Say that a functional $h : \mathcal{X} \to \mathbf{R}$ with $\pi(|h|) < \infty$ (where $\pi(f) = \int_{\mathcal{X}} f(x)\,\pi(dx)$) satisfies a usual central limit theorem (CLT) for a Markov chain $P$ if as $n \to \infty$, the distribution of $n^{-1/2} \sum_{i=1}^{n} [h(X_i) - \pi(h)]$ converges weakly to $N(0, v)$, where (with $\{X_n\}$ in stationarity)

$$v \;=\; \mathbf{Var}(X_0) + 2 \sum_{i=1}^{\infty} \mathbf{Cov}(X_0, X_i) \;<\; \infty\,. \tag{3}$$

(We say "usual" to distinguish this convergence from e.g. convergence to other distributions, or other normalisations besides $n^{-1/2}$; see also [7], [3], [27], [37].)

It is known ([12], Theorem 18.5.3; see also [3], [10]) that if $P$ is geometrically ergodic, then $h$ satisfies a usual CLT provided $\pi(|h|^{2+\delta}) < \infty$ for some $\delta > 0$. It was proven in [28], following [16], that if $P$ is geometrically ergodic and reversible, then $h$ satisfies a usual CLT whenever $\pi(h^2) < \infty$. However, geometric ergodicity is an overly strong assumption; for example, periodic Markov chains can never be geometrically ergodic but they can still satisfy CLTs.

The following theorem shows that, for reversible Markov chains, variance bounding is the "right" definition for CLTs, i.e. variance bounding (unlike geometric ergodicity) is the weakest property which still guarantees usual CLTs for all $L^2$ functionals. (We assume the stationary distribution for $P$ is unique, to avoid degenerate cases where the state space breaks up into multiple closed subsets.)

**Theorem 7.** *If $P$ is reversible, with unique stationary distribution $\pi(\cdot)$, then $P$ is variance bounding if and only if every $h : \mathcal{X} \to \mathbf{R}$ with $\pi(h^2) < \infty$ satisfies a usual CLT for $P$.*

**Remark.** There are other results available (see e.g. [3] and the references therein) which guarantee CLTs for *specific* functionals, rather than for all $L^2$ functionals. However, often

MCMC is used to generate samples from $\pi(\cdot)$ before it is decided which functionals are of statistical interest. Thus, we find that it is most useful having results like Theorem 7 which apply to all $L^2$ functionals simultaneously.

# 5. Relation to the Peskun ordering.

The following partial order on Markov chain kernels was introduced by Peskun [26] for finite state spaces, and later by Tierney [40] for general state spaces.

**Definition.** Let $P_1$ and $P_2$ be two Markov chain kernels on $(\mathcal{X}, \mathcal{F})$, both having invariant probability measure $\pi$. Then $P_1$ *dominates* $P_2$ *off the diagonal*, written $P_1 \succeq P_2$, if $P_1(x, A) \geq P_2(x, A)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{F}$ with $x \notin A$.

It was proved by Peskun [26] for finite state spaces, and then by Tierney [41] (see also [23], [22]) for general state spaces, that if $P_1 \succeq P_2$, and $P_1$ and $P_2$ are reversible with respect to the same $\pi(\cdot)$, then $\mathbf{Var}(h, P_1) \leq \mathbf{Var}(h, P_2)$ for all $h : \mathcal{X} \to \mathbf{R}$. That is, $P_1$ is "better" than $P_2$, in the sense of being uniformly more efficient for estimating expectations of functionals. Thus, it seems reasonable that any Markov chain property designed to indicate good estimation should be preserved under the Peskun ordering. For the variance bounding property, that is indeed the case:

**Theorem 8.** *If $P_1$ and $P_2$ are both reversible with respect to $\pi(\cdot)$, and $P_1 \succeq P_2$, and $P_2$ is variance bounding, then $P_1$ is variance bounding.*

On the other hand, the corresponding property for geometric ergodicity does not hold, indicating another advantage of variance bounding over geometric ergodicity:

**Example 9.** Let $\mathcal{X} = \mathbf{Z}$ with $\pi(m) = 2^{-|m|}\,/\,3$. Define $P_1$ by $P_1(x, x-1) = 2/3$ and $P_1(x, x+1) = 1/3$ for $x > 0$, and $P_1(x, x-1) = 1/3$ and $P_1(x, x+1) = 2/3$ for $x < 0$, and $P_1(0, -1) = P_1(0, 1) = 1/2$. Also, let $P_2$ be the Metropolis algorithm for $\pi(\cdot)$ with proposal distribution $Q(x, x+1) = Q(x, x-1) = 1/2$. (Thus, $P_2(x, x+1) = P_2(x, x) = 1/4$ and $P_2(x, x-1) = 1/2$ for $x > 0$; $P_2(x, x-1) = P_2(x, x) = 1/4$ and $P_2(x, x+1) = 1/2$ for $x < 0$; and $P_2(0, -1) = P_2(0, 1) = 1/4$ and $P_2(0, 0) = 1/2$.) Then both $P_1$ and $P_2$ are reversible with respect to $\pi(\cdot)$, and also $P_1 \succeq P_2$. Furthermore, it follows as in Mengersen and Tweedie [19] that $P_2$ is geometrically ergodic, and hence variance bounding by Theorem 1. On the other hand, $P_1$ is periodic, and hence cannot be geometrically ergodic, even though $P_1 \succeq P_2$. (Of course, $P_1$ is still variance bounding, by Theorem 8.)

## 6. Application to Metropolis-Hastings algorithms.

We now consider Metropolis-Hastings algorithms ([20], [9]). We define a slight generalisation, as follows. Given a reference measure $\nu(\cdot)$ on $\mathcal{X}$, with respect to which $\pi(dx) = t(x)\,\nu(dx)$, and a non-negative (measurable) function $q : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ with $\int_{\mathcal{X}} q(x,y)\,\nu(dy) \leq 1$ for all $x \in \mathcal{X}$, the *sub-Metropolis-Hastings algorithm* is the algorithm with transition kernel

$$M_q(x, dy) \;=\; \alpha(x,y)\, q(x,y)\, \nu(dy) \;+\; r(x)\, \delta_x(dy)\,,$$

where $\alpha(x,y) = \min\left(1,\; \frac{t(y)\, q(y,x)}{t(x)\, q(x,y)}\right)$, and $r(x) = 1 - \int_{\mathcal{X}} \alpha(x,y)\, q(x,y)\, \nu(dy) \geq 0$.

By construction, this algorithm is reversible with respect to $\pi(\cdot)$. It may be described as follows. With probability $\int_{\mathcal{X}} q(x,y)\,\nu(dy)$, it performs the usual Metropolis-Hastings algorithm with proposal density $q(x,y) \,/\, \int_{\mathcal{X}} q(x,y)\,\nu(dy)$. Otherwise, with probability $1 - \int_{\mathcal{X}} q(x,y)\,\nu(dy)$, it stays at its current state. If $\int_{\mathcal{X}} q(x,y)\,\nu(dy) = 1$, then $M_q$ is the usual Metropolis-Hastings algorithm.

By direct inspection, noting that $\alpha(x,y)\, q(x,y) = \min\left(q(x,y),\; \frac{t(y)}{t(x)}\, q(y,x)\right)$, we see the following:

**Proposition 10.** *For fixed $\nu(\cdot)$ and $t$, if $q_1(x,y) \geq q_2(x,y)$ for all $x, y \in \mathcal{X}$ with $x \neq y$, then $M_{q_1} \succeq M_{q_2}$. (Hence, by Theorem 8, if $M_{q_2}$ is variance bounding then so is $M_{q_1}$.)*

Now, suppose $M_{q_2}$ is variance bounding, and that $q_1(x,y) \geq c\, q_2(x,y)$ for all $x \neq y$, for some $c > 0$. We can assume (by replacing $c$ with $\min(c, 1)$ if necessary) that $c \leq 1$. Then $M_{cq_2} = c\, M_{q_2} + (1-c)I$. Hence, by Corollary 3 (with $a = 1 - c$), $M_{cq_2}$ is also variance bounding. It then follows from Proposition 10 that $M_{q_1}$ is also variance bounding. We conclude:

**Corollary 11.** *If $q_1(x,y) \geq c\, q_2(x,y)$ for all $x, y \in \mathcal{X}$ with $x \neq y$, for some $c > 0$, and if $M_{q_2}$ is variance bounding, then $M_{q_1}$ is variance bounding.*

Example 9 above shows that the analogous statement to Corollary 11 for geometric ergodicity does not hold.

To continue, call a (measurable) function $s : \mathcal{X} \to [0, \infty)$ *MT-good* if it is symmetric, positive, and continuous, with exponentially bounded tails, and with $\int_{-\infty}^{\infty} s(u)\, du = 1$. Then a result of Mengersen and Tweedie [19] (see also [32] for higher-dimensional analogs) says that a random-walk Metropolis algorithm on $\mathcal{X} = \mathbf{R}$, with proposal density $q(x,y) = s(y-x)$ for some MT-good $s$, is geometrically ergodic provided the target density has exponentially bounded tails. This is a very impressive result, but with the severe restriction that the proposal increments must correspond to a symmetric random walk. To improve this, we make the following definition.

**Definition.** A proposal density function $q : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ is a *uniformly minorised increment distribution (UMID)* if there is $c > 0$ and MT-good $s : \mathcal{X} \to [0, \infty)$ such that $q(x, y) \geq c\, s(y - x)$ for all $x, y \in \mathcal{X}$.

Combining Theorem 1 and Corollary 11 with the result of [19] immediately gives:

**Corollary 12.** *Let $t$ be a target density with exponentially bounded tails, and let $q$ be a UMID proposal density function. Then $M_q$ is variance bounding.*

Note that in Corollary 12, we do not need to assume that $s$ has exponentially bounded tails, since if not then we can simply replace $s(x)$ by $\min\left(s(x), e^{-|x|}\right)$ without affecting the conclusion. Note also that Corollary 12 does not require the proposal density $q$ to be symmetric, nor to correspond to a random walk. (Similar generalisations are also available for the multidimensional case, as in [32].)

As one application of Corollary 12, consider a Langevin (MALA) algorithm (see [33]), with proposal density given by $Q(x, \cdot) = N(x + \frac{1}{2}\,\delta\,\nabla \log t(x),\ \delta^2)$ for some $\delta > 0$. Now, if the target density $t$ is $C^1$ with tails that are precisely exponential, then $\nabla \log t(x)$ is a bounded function of $x \in \mathcal{X}$, and it follows easily that $q$ is UMID. We conclude:

**Corollary 13.** *A Langevin algorithm for a $C^1$ target density on $\mathcal{X} = \mathbf{R}$ with exponential tails is variance bounding.*

As a final application, we consider a Metropolis-Hastings algorithm for a density $t$ supported on $(0, \infty)$, with proposal distribution given by $Q(x, \cdot) = N(x, x^b)$ for some fixed $b > 0$. That is, the variance of the proposal increment depends on the current state $x \in \mathcal{X}$. (Related models were considered in [31].)

If $b > 2$, then as $x \to \infty$ the proposal values will be farther and farther out in the tails, so $\lim_{x\to\infty} P(x, \{x\}) = 1$. It follows as in [32], or by a simple capacitance argument (e.g. [17]), that the resulting Markov chain is neither geometrically ergodic nor variance bounding. So, we do not consider that case further here. (On the other hand, numerical simulations related to [31] indicate that if $t$ is e.g. a Cauchy distribution, then values $b \approx 2.7$ may give fastest numerical convergence, which is a separate but related issue.)

If $b = 2$, then the distribution $Q(x, \cdot)$ equals the distribution of $x + x\,Z$, where $Z \sim N(0, 1)$. Taking logarithms (cf. [31]) gives rise to an equivalent chain which is an ordinary random-walk Metropolis algorithm, with modified target density $\tilde{t}(y) = e^y\, t(e^y)$, and with increment density $f(u)$ equal to the density of $\log(1+Z)$ where $Z \sim N(0,1)$. This increment distribution is clearly UMID; indeed, we can simply let $c\,s(u) = \min\left(f(u), f(-u)\right)$. Hence, by Corollary 12, the transformed chain – and hence also the original chain – is variance bounding provided that $\tilde{t}$ has exponentially bounded tails.

Finally, suppose that $0 < b < 2$. Then $Q(x,\cdot)$ is the distribution of $x + x^{b/2} Z$, where $Z \sim N(0,1)$. Instead of logarithms, consider the transformation $X \mapsto X^a$, where $a = 1-b/2$ (so $0 < a < 1$). Then the proposal increment from $x \in \mathcal{X}$ transforms from $x^{b/2} Z$ to $W = h(Z) \equiv [x + x^{b/2} Z]^a - x^a = x^a(1 + x^{-a} Z)^a - x^a$. Inverting this, $Z = h^{-1}(W) = x^a((1 + W x^{-a})^{1/a} - 1)$. Now, the density of $Z$ is $g(z) = (2\pi)^{-1/2} e^{-z^2/2}$. Hence, for the transformed chain, the proposal increment $W$ has density

$$\frac{g\Big(h^{-1}(w)\Big)}{(dw/dz)} \;=\; \frac{e^{-h^{-1}(w)^2/2}}{\sqrt{2\pi}\, a\, (1 + x^{-a} z)^{a-1}}\,.$$

We compute that, as $x \to \infty$, this expression converges to $(2\pi)^{-1/2} a^{-1} e^{-(w/a)^2/2}$, i.e. to the density function of the $N(0, a^2)$ distribution. (Intuitively, this is because $(\frac{d}{dx} x^a)^2 x^b = a^2 x^{2a-2} x^b = a^2$ is constant, so the increment variance of the transformed chain is approximately stabilised.) Hence, for large enough $x$, and thus for all $x$ by positivity and continuity, the proposal density is UMID. Therefore, by Corollary 12, the transformed and original chains are variance bounding provided that the transformed target density has exponentially bounded tails.

# 7. Spectra and Theorem Proofs.

We now proceed to the proofs of the theorems. We begin by recalling some standard notation. Let $P$ be a Markov chain kernel with stationary distribution $\pi(\cdot)$ on a state space $(\mathcal{X}, \mathcal{F})$. For measurable $f, g : \mathcal{X} \to \mathbf{R}$, write $\langle f, g\rangle = \int_{\mathcal{X}} f(x)\, g(x)\, \pi(dx)$, and $\|f\| = \langle f, f\rangle^{1/2}$. Let $L_0^2(\pi) = \{f : \mathcal{X} \to \mathbf{R};\ \pi(f) = 0,\ \pi(f^2) < \infty\}$, and regard $P$ as an operator acting on $L_0^2(\pi)$, by $(Pf)(x) = \int_{\mathcal{X}} f(y)\, P(x, dy)$. Write $\sigma(P)$ for the spectrum of the operator $P$ acting on $L_0^2(\pi)$ (see e.g. [4], [36]). If $P$ is a reversible Markov chain, then $P$ is a self-adjoint operator with respect to $\langle\cdot,\cdot\rangle$, and also $\sigma(P) \subseteq [-1, 1]$ (cf. [7], [1]). Theorem 2 of [28] says that if $P$ is reversible, then $P$ is geometrically ergodic if and only if there is $r < 1$ with $\sigma(P) \subseteq [-r, r]$.

We have the following.

**Theorem 14.** *If $P$ is reversible, then $P$ is variance bounding if and only if* $\sup\left(\sigma(P)\right) < 1$.

**Proof.** Suppose first that $\sup(\sigma(P)) \equiv \Lambda < 1$. Then by Proposition 1 of [36], $\mathbf{Var}(h, P) \leq 2(1-\Lambda)^{-1}\mathbf{Var}_\pi(h)$ for all $h : \mathcal{X} \to \mathbf{R}$. Hence, $P$ is variance bounding with constant $K = 2(1-\Lambda)^{-1} < \infty$.

Conversely, suppose $\sup(\sigma(P)) = 1$. Let $\mathcal{E}$ be the spectral measure for $P$ (see e.g. [4], [38], [7], [28], [36]), and let $r < 1$. Then $\mathcal{E}((r, 1])$ is non-zero, so there is $h \in L_0^2(\pi)$ in range of $\mathcal{E}((r, 1])$. It follows similarly to Proposition 1 of [36] (cf. [16], [7]) that $\mathbf{Var}(h, P) \geq \frac{1}{1-r}\,\mathbf{Var}_\pi(h)$. Since this holds for any $r < 1$, it follows that $\sup_{h \in \mathcal{L}_0^2(\pi)} [\mathbf{Var}(h, P)\,/\,\mathbf{Var}_\pi(h)] \geq \sup_{r<1} \frac{1}{1-r} = \infty$. Hence, $P$ is not variance bounding. ∎

**Proof of Theorem 1.** If $P$ is reversible and geometrically ergodic, then there is $r < 1$ with $\sigma(P) \subseteq [-r, r]$. In particular, $\sup(\sigma(P)) \leq r < 1$, so $P$ is variance bounding by Theorem 14. ∎

**Proof of Theorem 2.** (i) $\Longrightarrow$ (ii): Suppose $P$ is variance bounding, and $0 < a < 1$. Then by Theorem 14, $\sup(\sigma(P)) < 1$, i.e. there is $c < 1$ with $\sigma(P) \subseteq [-1, c]$. On the other hand,

$$\begin{aligned}
\sigma(aI + (1-a)P) &= \{\lambda \in \mathbf{R}\ s.t.\ (aI + (1-a)P - \lambda I) \text{ is not invertible}\} \\
&= \Big\{\lambda \in \mathbf{R}\ s.t.\ (1-a)\Big(P - \frac{\lambda - a}{1-a}\ I\Big) \text{ is not invertible}\Big\} \\
&= \Big\{\lambda \in \mathbf{R}\ s.t.\ \frac{\lambda - a}{1-a} \in \sigma(P)\Big\} \\
&= \{(a + (1-a)y)\ s.t.\ y \in \sigma(P)\}\,,
\end{aligned}$$

where the last equality follows by solving for $\lambda$ in the equation $y = \frac{\lambda - a}{1-a}$. Hence, since $\sigma(P) \subseteq [-1, c]$, it follows that

$$\sigma(aI + (1-a)P) \ \subseteq\ [a + (1-a)(-1),\ a + (1-a)c] \ =\ [2a-1,\ a + (1-a)c] \ \subseteq\ [-r, r]\,,$$

where $r = \max(|2a-1|,\ a + (1-a)c) < 1$. Hence, by Theorem 2 of [28], $aI + (1-a)P$ is geometrically ergodic.

(ii) $\Longrightarrow$ (iii): Immediate.

(iii) $\Longrightarrow$ (i): If $aI + (1-a)P$ is geometrically ergodic, then there is $r < 1$ with $\sigma(aI + (1-a)P) \subseteq [-r, r]$. But from the above,

$$\sigma(aI + (1-a)P) \ =\ \Big\{\lambda \in \mathbf{R}\ s.t.\ \frac{\lambda - a}{1-a} \in \sigma(P)\Big\}$$

so it follows that $\sigma(P) \subseteq [\frac{-r-a}{1-a},\ \frac{r-a}{1-a}]$. In particular, $\sup(\sigma(P)) \leq \frac{r-a}{1-a} < 1$, so $P$ is variance bounding. ∎

**Proof of Corollary 3.** We see from the proof of Theorem 2 that

$$\sup\Big(\sigma(aI + (1-a)P)\Big) \;=\; a + (1-a)\,\sup\left(\sigma(P)\right).$$

It follows that for $0 \le a < 1$, $\sup\left(\sigma(aI + (1-a)P)\right) < 1$ if and only if $\sup\left(\sigma(P)\right) < 1$. The result then follows from Theorem 14. ■

**Proof of Theorem 4.** If $P$ is reversible and geometrically ergodic, then $P$ is variance bounding by Theorem 1. Conversely, suppose $P$ is reversible and variance bounding, with $\delta \equiv \inf_{x \in \mathcal{X}} P(x, \{x\}) > 0$. Let $S(x, A) = (1-\delta)^{-1}(P(x, A) - \delta\, \mathbf{1}_{x \in A})$. Then $S$ is another Markov chain kernel on $\mathcal{X}$, and $P = \delta\, I + (1-\delta) S$. It follows that $\inf \sigma(P) \ge \delta + (1-\delta)(-1) = 2\delta - 1 > -1$. Also $\sup \sigma(P) < 1$ by Theorem 14. Hence there is $r < 1$ with $\sigma(P) \subseteq [-r, r]$, so $P$ is geometrically ergodic. ■

**Proof of Theorem 5.** Note that $\mathbf{E}[f(X_0)\, f(X_1)] = \langle f,\, Pf \rangle$, so positivity is equivalent to $\langle f,\, Pf \rangle \ge 0$ for all $f \in L_0^2(\pi)$. This implies that $\lambda \ge 0$ for all $\lambda \in \sigma(P)$. Hence, using Theorem 14,

$$\begin{aligned} P \text{ is geometrically ergodic} \;&\Longleftrightarrow\; \sup\{|\lambda| : \lambda \in \sigma(P)\} \;<\; 1 \\ &\Longleftrightarrow\; \sup\{\lambda : \lambda \in \sigma(P)\} \;<\; 1 \\ &\Longleftrightarrow\; P \text{ is variance bounding}\,. \end{aligned}$$

■

**Proof of Theorem 6.** Note that we can write $\overline{P}^n = [\frac{1}{2}(I+P)]^n$. Hence, the result follows immediately from Theorem 2 (with $a = 1/2$). ■

**Proof of Theorem 7.** If $P$ is variance bounding, then $\Lambda \equiv \sup\left(\sigma(P)\right) < 1$. Let $\mathcal{E}$ be the spectral measure for $P$, and let $\mathcal{E}_h$ the induced measure defined by

$$\mathcal{E}_h(S) \;=\; \int_X h(x)\; (\mathcal{E}(S)h)(x)\; \pi(dx)\,.$$

Then it follows (cf. [7]) that

$$\sigma^2 \;\equiv\; \int_{-1}^{1} \frac{1+\lambda}{1-\lambda}\, \mathcal{E}_h(d\lambda) \;\le\; \frac{1+\Lambda}{1-\Lambda} \;<\; \infty\,.$$

It then follows from Kipnis and Varadhan [16] (see also [3]) that $h$ satisfies a usual CLT for $P$.

Conversely, if $P$ is not variance bounding, then $\Lambda = 1$. It follows as in the proof of Theorem 14 that $\mathcal{E}\Big((r,1]\Big)$ is non-zero for every $r < 1$. Since $P$ has unique stationary distribution, $1 \notin \sigma(P)$, so there must be infinitely many $m \in \mathbf{N}$ such that $\mathcal{E}\Big((1-2^{-m},\ 1-2^{-m-1}]\Big)$ is non-zero. Let $m_1 < m_2 < \ldots$ (so $m_i \geq i$) with $\mathcal{E}\Big((1-2^{-m_i},\ 1-2^{-m_i-1}]\Big)$ non-zero. Let $g_i \in L^2_0(\pi)$ be in the range of $\mathcal{E}\Big((1-2^{-m_i},\ 1-2^{-m_i-1}]\Big)$, with $\|g_i\| = 1$. Then spectral theory implies that the $\{g_i\}$ are orthonormal, and furthermore $\mathbf{Cov}(g_i,\ Pg_i) = \langle g_i,\ Pg_i\rangle \geq 1-2^{-m_i}$. Finally, let $h = \sum_{i=1}^{\infty} 2^{-i/2} g_i$. Then by orthonormality,

$$\mathbf{Var}_\pi(h) \ = \ \|h\|^2 \ = \ \sum_{i=1}^{\infty}\Big(2^{-i/2}\Big)^2 \ = \ 1 \ < \ \infty\,.$$

On the other hand, with $\{X_n\}$ in stationarity, again using orthonormality,

$$\begin{aligned}\mathbf{Cov}(h(X_0), h(X_n)) \ &= \ \sum_{i=1}^{\infty} 2^{-i}\,\mathbf{Cov}(g_i,\ P^n g_i) \\ &\geq \ \sum_{i=1}^{\infty} 2^{-i}(1-2^{-m_i})^n \ \geq \ \sum_{i=1}^{\infty} 2^{-i}(1-2^{-i})^n\,.\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{n=0}^{\infty}\mathbf{Cov}(h(X_0), h(X_n)) \ &\geq \ \sum_{i=1}^{\infty} 2^{-i}\sum_{n=0}^{\infty}(1-2^{-i})^n \\ &= \ \sum_{i=1}^{\infty} 2^{-i}[1-(1-2^{-i})]^{-1} \ = \ \sum_{i=1}^{\infty} 2^{-i}\,2^{i} \ = \ \sum_{i=1}^{\infty}(1) \ = \ \infty\,.\end{aligned}$$

It follows that $v$ in (3) is infinite, so $h$ does not satisfy a usual CLT for $P$. ∎

**Proof of Theorem 8.** Lemma 3 of Tierney [41] says that since $P_1 \succeq P_2$, therefore $P_2 - P_1$ is a positive operator. It follows that $\sup(\sigma(P_2)) \geq \sup(\sigma(P_1))$. Hence, using Theorem 14 twice, if $P_2$ is variance bounding, then $\sup(\sigma(P_2)) < 1$, so $\sup(\sigma(P_1)) < 1$, so $P_1$ is variance bounding. (Alternatively, by Theorem 4 of [41], $\mathbf{Var}(h, P_1) \leq \mathbf{Var}(h, P_2) \leq K\,\mathbf{Var}_\pi(h)$.) ∎

**Remark.** The above theorems have all been proven for reversible chains only. However, it seems likely that analogs of some of them (e.g. Theorem 1) carry over in some form to non-reversible chains, about which various facts about convergence are known (see e.g. [6], [18], [5], [23]). We leave this as an open problem for future work.

# 8. Summary.

This paper defined a Markov chain to be variance bounding if the asymptotic variances for functionals with unit stationary variance are uniformly bounded. For reversible chains, we proved that this property is weaker than geometric ergodicity, but equivalent to $aI + (1 - a)P$ being geometrically ergodic for all $0 < a < 1$. Furthermore, in contrast to geometric ergodicity, the variance bounding property: allows for periodicity; is equivalent to all $L^2$ functionals satisfying a usual central limit theorem; and is preserved under the Peskun [26] partial ordering on Markov chains. We also presented some applications to Metropolis-Hastings MCMC algorithms, and showed how variance bounding could apply more easily and more generally than geometric ergodicity.

Overall, we view these results as indicating that as a property to use in the study of MCMC algorithms, variance bounding is similar to, but more convenient than, geometric ergodicity. We hope that the notion of variance bounding can be used to further understand Markov chains and MCMC algorithms in other contexts.

**Acknowledgements.** We thank the anonymous referee for many very helpful comments.

# References

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

[2] R.C. Bradley (2005), Basic properties of strong mixing conditions: a survey and some open questions. Prob. Surv. **2**, 107–144.

[3] K.S. Chan and C.J. Geyer (1994), Discussion to reference [40]. Ann. Stat. **22**, 1747–1758.

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

[5] P. Diaconis, S. Holmes, and R.M. Neal, Analysis of a non-reversible Markov chain sampler. Ann. Appl. Prob. **10**, 726–752.

[6] J.A. Fill (1991), Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with an application to the exclusion process. Ann. Appl. Prob. **1**, 62–87.

[7] C.J. Geyer (1992), Practical Markov chain Monte Carlo. Stat. Sci. **7**, 473-483.

[8] W.R. Gilks and G.O. Roberts (1995), Strategies for improving MCMC.R. Gilks). In *MCMC in Practice* (eds. W.R. Gilks, D.J. Spiegelhalter and S. Richardson), Chapman and Hall, 89–114.

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

[10] J.P. Hobert, G.L. Jones, B. Presnell, and J.S. Rosenthal (2002), On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo. Biometrika **89**, 731–743.

[11] J.P. Hobert and J.S. Rosenthal (2007), Norm comparisons for data augmentation. Preprint.

[12] I.A. Ibragimov and Y.V. Linnik (1971), Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen (English translation).

[13] G.L. Jones (2004), On the Markov chain central limit theorem. Prob. Surv. **1** (2004), 299–320.

[14] G.L. Jones and J.P. Hobert (2001), Honest exploration of intractable probability distributions via Markov chain Monte Carlo. Stat. Sci. **16**, 312–334.

[15] G.L. Jones and J.P. Hobert (2004), Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. Ann. Stat. **32**, 784–817.

[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] G.F. Lawler and A.D. Sokal (1988), Bounds on the $L^2$ spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. Trans. Amer. Math. Soc. **309**, 557–580.

[18] J.S. Liu, W. Wong, and A. Kong (1994), Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika **81**, 27-40.

[19] K.L. Mengersen and R.L. Tweedie (1996), Rates of convergence of the Hastings and Metropolis algorithms. Ann. Stat. **24**, 101–121.

[20] 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.

[21] S.P. Meyn and R.L. Tweedie (1993), Markov chains and stochastic stability. Springer-Verlag, London. Available at probability.ca/MT.

[22] A. Mira (2001), Ordering and improving the performance of Monte Carlo Markov chains. Stat. Sci. **16**, 340–350.

[23] A. Mira and C.J. Geyer (2000), On non-reversible Markov chains. In N. Madras (ed.), Fields Institute Communications, Volume **26**: Monte Carlo Methods, pp. 93–108. Providence, RI: American Mathematical Society.

[24] A. Mira, J. Mller and G. O. Roberts (2001), Perfect Slice Samplers. J. Royal Stat. Soc. Series B **63**, 593–606.

[25] R.M. Neal (2003), Slice sampling (with discussion). Ann. Stat. **31**, 705–767.

[26] P.H. Peskun (1973), Optimum Monte Carlo sampling using Markov chains. Biometrika **60**, 607–612.

[27] G.O. Roberts (1999), A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. J. Appl. Prob. **36**, 1210–1217.

[28] G.O. Roberts and J.S. Rosenthal (1997), Geometric ergodicity and hybrid Markov chains. Electronic Comm. Prob. **2**, Paper no. 2, 13–25.

[29] G.O. Roberts and J.S. Rosenthal (1998), Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). Canadian J. Stat. **26**, 5–31.

[30] G.O. Roberts and J.S. Rosenthal (1999), Convergence of slice sampler Markov chains. J. Royal Stat. Soc. Series B **61**, 643–660.

[31] G.O. Roberts and J.S. Rosenthal (2006), Examples of Adaptive MCMC. Preprint.

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

[33] G. O. Roberts and R. L. Tweedie (1996), Exponential convergence of Langevin diffusions and their discrete approximations. Bernoulli **2**, 341–364.

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

[35] J.S. Rosenthal (2002), Quantitative convergence rates of Markov chains: A simple account. Electronic Comm. Prob. **7**, 123–128.

[36] J.S. Rosenthal (2003), Asymptotic variance and convergence rates of nearly-periodic MCMC algorithms. J. Amer. Stat. Assoc. **98**, 169–177.

[37] O. Häggström and J.S. Rosenthal (2007), On variance conditions for Markov chain CLTs. Preprint.

[38] W. Rudin (1991), Functional Analysis, 2nd ed. McGraw-Hill, New York.

[39] 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.

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

[41] L. Tierney (1998), A note on Metropolis-Hastings kernels for general state spaces. Ann. Appl. Prob. **8**, 1–9.