Probability Surveys
Vol. 1 (2004) 299–320
ISSN: 1549-5787
DOI: 10.1214/154957804100000051

On the Markov chain central limit theorem*

Galin L. Jones

School of Statistics,
University of Minnesota,
Minneapolis, MN, USA
e-mail: galin@stat.umn.edu

Abstract: The goal of this expository paper is to describe conditions which guarantee a central limit theorem for functionals of general state space Markov chains. This is done with a view towards Markov chain Monte Carlo settings and hence the focus is on the connections between drift and mixing conditions and their implications. In particular, we consider three commonly cited central limit theorems and discuss their relationship to classical results for mixing processes. Several motivating examples are given which range from toy one-dimensional settings to complicated settings encountered in Markov chain Monte Carlo.

Keywords and phrases: Central Limit Theorem, Markov Chain, Monte Carlo, Mixing Condition, Drift Condition.

Received September 2004.

1. Introduction

Let $X = \{X_i : i = 0, 1, 2, \dots\}$ be a Harris ergodic Markov chain on a general space X with invariant probability distribution π having support X. Let f be a Borel function and define $\bar{f}_n := n^{-1} \sum_{i=1}^{n} f(X_i)$ and $\mathrm{E}_\pi f := \int_{\mathsf{X}} f(x)\pi(dx)$. When $\mathrm{E}_\pi|f| < \infty$ the ergodic theorem guarantees that $\bar{f}_n \to \mathrm{E}_\pi f$ with probability 1 as $n \to \infty$. The main goal here is to describe conditions on X and f under which a central limit theorem (CLT) holds for $\bar{f}_n$; that is,

$$\sqrt{n}(\bar{f}_n - \mathrm{E}_\pi f) \xrightarrow{d} \mathrm{N}(0, \sigma_f^2) \tag{1}$$

as $n \to \infty$ where $\sigma_f^2 := \mathrm{var}_\pi\{f(X_0)\} + 2\sum_{i=1}^{\infty} \mathrm{cov}_\pi\{f(X_0), f(X_i)\} < \infty$. Although all of the results presented in this paper hold more generally, the primary motivation is found in Markov chain Monte Carlo (MCMC) settings where the existence of a CLT is an extremely important practical problem. Often π is high dimensional or known only up to a normalizing constant but the value of $\mathrm{E}_\pi f$ is required. If X can be simulated then $\bar{f}_n$ is a natural estimate of $\mathrm{E}_\pi f$. The existence of a CLT then allows one to estimate σ_f^2 in order to decide if $\bar{f}_n$ is a good estimate of $\mathrm{E}_\pi f$. (Estimation of σ_f^2 is challenging and requires specialized techniques that will not be considered further here; see [32] and [21] for an introduction.) Thus the existence of a CLT is crucial to sensible implementation

*This is an original survey paper.

of MCMC; see [33] for more on this point of view. The following simple example illustrates one of the situations common in MCMC settings.

Example 1. Consider a simple hard-shell (also known as hard-core) model. Suppose $\mathcal{X} = \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \subseteq \mathbb{Z}^2$. A *proper* configuration on $\mathcal{X}$ consists of coloring each point either black or white in such a way that no two adjacent points are white. Let X denote the set of all proper configurations on $\mathcal{X}$, $N_{\mathsf{X}}(n_1, n_2)$ be the total number of proper configurations and π be the uniform distribution on X so that each proper configuration is equally likely. Suppose our goal is to calculate the typical number of white points in a proper configuration; that is, if $W(x)$ is the number of white points in $x \in \mathsf{X}$ then we want the value of

$$\mathrm{E}_\pi W = \sum_{x \in \mathsf{X}} \frac{w(x)}{N_{\mathsf{X}}(n_1, n_2)} \ .$$

If n_1 and n_2 are even moderately large then we will have to resort to an approximation to $\mathrm{E}_\pi W$. Consider the following Markov chain on X. Fix $p \in (0, 1)$ and set $X_0 = x_0$ where $x_0 \in \mathsf{X}$ is an arbitrary proper configuration. Randomly choose a point $(x, y) \in \mathcal{X}$ and independently draw $U \sim \text{Uniform}(0, 1)$. If $u \leq p$ and all of the adjacent points are black then color (x, y) white leaving all other points alone. Otherwise, color (x, y) black and leave all other points alone. Call the resulting configuration X_1. Continuing in this fashion yields a Harris ergodic Markov chain $\{X_0, X_1, X_2, \ldots\}$ having π as its invariant distribution. It is now a simple matter to estimate $\mathrm{E}_\pi W$ with $\bar{w}_n$. Also, since X is finite (albeit potentially large) it is well known that X will converge exponentially fast to π which implies that a CLT holds for $\bar{w}_n$.

Following the publication of Meyn and Tweedie's influential book [41] the use of drift and minorization conditions has become a popular method for establishing the existence of a CLT. Indeed without this constructive methodology it is difficult to envision how one would deal with complicated situations encountered in MCMC. In turn, this has led much of the recent work on general state space Markov chains to focus on the implications of drift and minorization. Another outcome of this approach is that classical results in mixing processes have been somewhat neglected. For example, two recent reviews of Markov chain theory and its connection to MCMC ([47], [57]) consider CLTs for Markov chains but neither contains any substantive discussion of the results from mixing processes. On the other hand, work on mixing processes rarely discusses their applicability to the important Markov chain setting outside of the occasional discrete state space example. For example, in [5] there is a recommended review of CLTs for mixing processes but no mention of the connections with Markov chains. Also, Robert gave a brief discussion of the implication of mixing conditions for Markov chain CLTs in [50] but failed to connect them to the use of drift conditions. Thus one of the main goals of this article is to consider the connections between drift and minorization and mixing conditions and their implications for the CLT for general state space Markov chains.

2. Markov Chains and Examples

Let $P(x, dy)$ be a Markov transition kernel on a general space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and write the associated discrete time Markov chain as $X = \{X_i : i = 0, 1, 2, \ldots\}$. For $n \in \mathbb{N} := \{1, 2, 3, \ldots\}$, let $P^n(x, dy)$ denote the n-step Markov transition kernel corresponding to P. Then for $i \in \mathbb{N}$, $x \in \mathsf{X}$ and a measurable set A, $P^n(x, A) = \Pr(X_{n+i} \in A | X_i = x)$. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function and define $Pf(x) := \int f(y)P(x, dy)$ and $\Delta f(x) := Pf(x) - f(x)$. Always, X will be assumed to be Harris ergodic, that is, aperiodic, ψ-irreducible and positive Harris recurrent; for definitions see [41] or [46]. These assumptions are more than enough to guarantee a strong form of convergence: For every initial probability measure $\lambda(\cdot)$ on $\mathcal{B}(\mathsf{X})$

$$\|P^n(\lambda, \cdot) - \pi(\cdot)\| \to 0 \quad \text{as} \quad n \to \infty \tag{2}$$

where $P^n(\lambda, A) := \int_{\mathsf{X}} P^n(x, A)\lambda(dx)$ and $\| \cdot \|$ is the total variation norm. Throughout we will be concerned with the rate of this convergence. Let $M(x)$ be a nonnegative function and $\gamma(n)$ be a nonnegative decreasing function on $\mathbb{Z}_+$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \le M(x)\gamma(n) \; . \tag{3}$$

When X is *geometrically ergodic* (3) holds with $\gamma(n) = t^n$ for some $t < 1$. *Uniform ergodicity* means M is bounded and $\gamma(n) = t^n$ for some $t < 1$. *Polynomial ergodicity of order m* where $m \ge 0$ corresponds to $\gamma(n) = n^{-m}$.

Establishing (3) directly may be difficult when X is a general space. However, some constructive methods are given in the following brief discussion; the interested reader should consult [31] and [41] for a more complete introduction to these methods.

A *minorization condition* holds on a set C if there exists a probability measure Q on $\mathcal{B}(\mathsf{X})$, a positive integer n_0 and an $\epsilon > 0$ such that

$$P^{n_0}(x, A) \ge \epsilon\, Q(A) \quad \forall\, x \in C \; , \; A \in \mathcal{B}(\mathsf{X}) \; . \tag{4}$$

In this case, C is said to be *small.* If (4) holds with $C = \mathsf{X}$ then X is uniformly ergodic and, as is well-known,

$$\|P^n(x, \cdot) - \pi(\cdot)\| \le (1 - \epsilon)^{\lfloor n/n_0 \rfloor} \; .$$

Uniformly ergodic Markov chains are rarely encountered in MCMC unless X is finite or bounded.

Geometric ergodicity may be established via the following drift condition: Suppose that for a function $V : \mathsf{X} \to [1, \infty)$ there exist constants $d > 0$, $b < \infty$ such that

$$\Delta V(x) \le -dV(x) + bI(x \in C) \qquad x \in \mathsf{X} \tag{5}$$

where C is a small set and I is the usual indicator function.

Polynomial ergodicity may be established via a slightly different drift condition: Suppose that for a function $V : \mathsf{X} \to [1, \infty)$ there exist constants $d > 0$, $b < \infty$ and $0 \le \tau < 1$ such that

$$\Delta V(x) \le -d[V(x)]^\tau + bI(x \in C) \qquad x \in \mathsf{X} \tag{6}$$

where C is a small set. In [31] it is shown that (6) implies that X is polynomially ergodic of degree $\tau/(1-\tau)$. Recently, this drift condition has been generalized to other subgeometric (slower than geometric) rates of convergence in [16].

Remark 1. Either of the drift conditions (5) or (6) imply that in (3) we can take $M(x) \propto V(x)$. Moreover, Theorem 14.3.7 in [41] shows that if (5) holds then $\mathrm{E}_\pi V < \infty$. Since the stronger property of V-uniform ergodicity is equivalent to (5) [41, Chapter 16] we conclude that geometrically (and uniformly) ergodic Markov chains satisfy (3) with $\mathrm{E}_\pi M < \infty$ (Also, see Fact 10 in [57]). On the other hand, the polynomial drift (6) only seems to imply that $\mathrm{E}_\pi V^\tau < \infty$ where $\tau < 1$. Thus, when (6) holds, one way to ensure that $\mathrm{E}_\pi M < \infty$ is to show that $\mathrm{E}_\pi V < \infty$.

Beyond establishing a rate of convergence, drift conditions also immediately imply the existence of a CLT for certain functions.

Theorem 1. *Let X be a Harris ergodic Markov chain on X having stationary distribution π. Suppose $f : \mathsf{X} \to \mathbb{R}$ and assume that one of the following conditions hold:*

1. *The drift condition (5) holds and $f^2(x) \le V(x)$ for all $x \in \mathsf{X}$.*
2. *The drift condition (6) holds and $|f(x)| \le V(x)^{\tau+\eta-1}$ for all $x \in \mathsf{X}$ where $1-\tau \le \eta \le 1$ is such that $E_\pi V^{2\eta} < \infty$.*

Then $\sigma_f^2 \in [0, \infty)$ and if $\sigma_f^2 > 0$ then for any initial distribution

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2)$$

as $n \to \infty$.

Remark 2. The first part of the theorem is Theorem 17.0.1 in [41] while the second part is Theorem 4.2 in [31].

Remark 3. The rate of convergence in the CLT when the drift condition (5) holds has been investigated in [36].

There has been a substantial amount of effort devoted to establishing drift and minorization conditions in MCMC settings. For example, Gibbs samplers were examined in [26], [34], [39], [50], [53], [60, 61] and [62]. While Metropolis-Hastings-Green (MHG) algorithms were considered in [8], [16], [18], [19], [22], [30], [31], [42], and [40]. Also, slice samplers were analyzed in [43] and [55].

In the next section three simple examples are presented in order to give the reader a taste of using these results in specific models and to demonstrate the application of Theorem 1. More substantial examples will be considered in Section 5.

2.1. Examples

Example 1 continued. Since X is finite it is easy to see that (4) holds with $C = \mathsf{X}$ and hence the Markov chain described in Example 1 is uniformly ergodic. Of course, if n_1 and n_2 are reasonably large ϵ may be too small to be useful.

Example 2. Suppose X lives on $\mathsf{X} = \mathbb{Z}$ such that if $x \geq 1$ and $0 < \theta < 1$ then

$$P(x, x+1) = P(-x, -x-1) = \theta \,, \qquad P(x, 0) = P(-x, 0) = 1 - \theta \,,$$

$$P(0, 1) = P(0, -1) = \frac{1}{2} \,.$$

This chain is Harris ergodic and has stationary distribution given by $\pi(0) = (1-\theta)/(2-\theta)$ and for $x \geq 1$

$$\pi(x) = \pi(-x) = \pi(0) \frac{\theta^{x-1}}{2} \,.$$

In Appendix A the drift condition (5) is verified with $V(x) = a^{|x|}$ for $a > 1$ satisfying $a\theta < 1$ and $(a\theta - 1)a + 1 - \theta < 0$ and $C = \{0\}$. Hence a CLT holds for $\bar{f}_n$ if $f^2(x) \leq a^{|x|} \; \forall \, x \in \mathbb{Z}$.

Example 3. A polynomial rate of convergence is established in both [31] and [63] for the random walk on $[0, \infty)$ determined by

$$X_{n+1} = (X_n + W_{n+1})^+$$

where $W_1, W_2, \ldots$ is a sequence of independent and identically distributed real-valued random variables. As long as $\mathrm{E}(W_1) < 0$ this chain will be Harris ergodic. When $\mathrm{E}(W_1^+)^m < \infty$ for some $m \geq 2$ the drift condition (6) is established in [31] with $V(x) = (x+1)^m$, $\tau = (m-1)/m$ and $C = [0, k]$ for some $k < \infty$. Hence a CLT holds for $\bar{f}_n$ if $|f(x)| \leq (x+1)^{m(\tau+\eta-1)}$ for all $x \geq 0$ where $1 - \tau \leq \eta \leq 1$ is such that $\mathrm{E}_\pi (x+1)^{2m\eta} < \infty$. Note that this moment condition also implies that $\mathrm{E}_\pi V < \infty$ as long as $\eta \geq 1/2$. Hence by an earlier remark $\mathrm{E}_\pi M < \infty$ with M as in (3).

Two things are clear: (i) drift and minorization provide powerful constructive tools for establishing a rate of convergence in total variation; and (ii) they are less impressive (but often useful!) tools for establishing CLTs in that the results in Theorem 1 depend on the non-unique function V.

3. Mixing Sequences

The goal of this section is to introduce three types of mixing conditions and discuss some of the connections with the total variation convergence in (2) and (3). There are a variety of mixing conditions (e.g. absolute regularity) that will not be considered here since they don't seem to have much impact on the CLT. Roughly speaking, mixing conditions are all attempts to quantify the rate at which events in the distant future become independent of the past.

Let $Y := \{Y_n\}$ denote a general sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and let $\mathcal{F}_k^m = \sigma(Y_k, \ldots, Y_m)$.

Definition 1. The sequence Y is said to be strongly mixing (or α–mixing) if $\alpha(n) \to 0$ as $n \to \infty$ where

$$\alpha(n) := \sup_{k \geq 1} \sup_{A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^\infty} |\mathcal{P}(A \cap B) - \mathcal{P}(A)\mathcal{P}(B)| .$$

Harris ergodic Markov chains are strongly mixing. Recall the coupling inequality [37, p. 12]:

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq \Pr(T > n) \tag{7}$$

where T is the usual coupling time of two Markov chains; one started in stationarity and one started arbitrarily. Under our assumptions the coupling time is almost surely finite and $\Pr(T > n) \to 0$ as $n \to \infty$. Let A and B be Borel sets so that by (7)

$$|P^n(x, A) - \pi(A)| \leq \Pr(T > n)$$

and

$$\begin{aligned} \Pr(T > n) &\geq \int_B |P^n(x, A) - \pi(A)|\pi(dx) \\ &\geq |\int_B [P^n(x, A) - \pi(A)]\pi(dx)| \\ &= |\Pr(X_n \in A \text{ and } X_0 \in B) - \pi(A)\pi(B)| . \end{aligned}$$

Then $\alpha(n) \leq \Pr(T > n)$ and hence $\alpha(n) \to 0$ as $n \to \infty$. Moreover, the rate of total variation convergence bounds the rate of α–mixing: If (3) holds with $\mathrm{E}_\pi M < \infty$, a similar argument shows that $\alpha(n) \leq \gamma(n)\mathrm{E}_\pi M$ and hence $\alpha(n) = O(\gamma(n))$. For example, geometrically ergodic Markov chains enjoy exponentially fast strong mixing.

Suppose the process Y is strictly stationary and let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function. Define the process $W := \{W_n = f(Y_n)\}$. Set $\mathcal{G}_k^m := \sigma(W_k, \ldots, W_m)$; hence $\mathcal{G}_k^m \subseteq \mathcal{F}_k^m$. Let α_W and α_Y be the strong mixing coefficients for the processes W and Y, respectively. Then $\alpha_W(n) \leq \alpha_Y(n)$. Similar comments apply to the mixing conditions given below. This elementary observation is fundamental to the proofs of the Markov chain CLTs considered in the sequel.

Definition 2. The sequence Y is said to be asymptotically uncorrelated (or ρ–mixing) if $\rho(n) \to 0$ as $n \to \infty$ where

$$\rho(n) := \sup\{\text{corr}(U, V), \, U \in L_2(\mathcal{F}_1^k) \, , V \in L_2(\mathcal{F}_{k+n}^\infty) \, k \geq 1\}.$$

It is standard that ρ-mixing sequences are also strongly mixing and, in fact, $4\alpha(n) \leq \rho(n)$. It is a consequence of the strong Markov property that if a Harris ergodic Markov chain is ρ-mixing then it enjoys exponentially fast ρ-mixing [4, Theorem 4.2] in the sense that there exists a $\theta > 0$ such that $\rho(n) = O(e^{-\theta n})$.

A necessary and sufficient condition for a Markov chain to be ρ–mixing is developed in [59] but before stating it a slight digression is required. Define the Hilbert space $L^2(\pi) := \{f \, : \mathsf{X} \to \mathbb{R} \, ; \mathrm{E}_\pi f^2 < \infty\}$ with inner product

$(f,g) = \mathrm{E}_\pi[f(x)g(x)]$ and norm $\|\cdot\|_2$. Let $L^2_0(\pi) := \{f \in L^2(\pi)\,;\, \mathrm{E}_\pi f = 0\}$ and note that if $f,g \in L^2_0(\pi)$ then $(f,g) = \mathrm{cov}_\pi(f,g)$. The kernel P defines an operator $T : L^2(\pi) \to L^2(\pi)$ via

$$(Tf)(x) = \int P(x,dy)f(y) \; .$$

It is easy to show that T is a contraction (i.e., $\|T\| \le 1$). Also, T is self-adjoint if and only if the kernel P satisfies *detailed balance* with respect to π:

$$\pi(dx)P(x,dy) = \pi(dy)P(y,dx) \qquad \forall\, x,y \in \mathsf{X} \; . \tag{8}$$

A Harris ergodic Markov chain is ρ–mixing if and only if

$$\lim_{n\to\infty} \sup_{\substack{f \in L^2_0(\pi) \\ \|f\|_2 = 1}} \|T^n f\|_2 = 0 \; . \tag{9}$$

There has been some work done on establishing sufficient conditions for Markov chains to be ρ-mixing. For example, in [38] it is shown that if the operator induced by a Gibbs sampler satisfies a Hilbert–Schmidt condition then it is ρ-mixing. However, the most interesting case is given in [54] whose Theorem 2.1 shows that if X is geometrically ergodic and (8) holds then there exists a $c < 1$ such that $\|Tf\|_2 \le c^2$ and $\|T^n f\|_2 = \|Tf\|_2^n$ hence (9) holds. We conclude that if X is geometrically ergodic and (8) holds then X is asymptotically uncorrelated.

Remark 4. Many Markov chains satisfy (8), indeed the MHG algorithm satisfies (8) by construction. However, (8) does not hold for those Markov chains associated with systematic scan Gibbs samplers and the Markov chain in Example 2, for example.

Definition 3. The sequence Y is said to be uniformly mixing (or ϕ–mixing) if $\phi(n) \to 0$ as $n \to \infty$ where

$$\phi(n) := \sup_{k \ge 1} \sup_{\substack{A \in \mathcal{F}_1^k, \mathcal{P}(A) \ne 0 \\ B \in \mathcal{F}_{k+n}^\infty}} |\mathcal{P}(B|A) - \mathcal{P}(B)| \; .$$

Uniformly mixing sequences are also asymptotically uncorrelated and strongly mixing. Moreover, $\rho(n) \le 2\sqrt{\phi(n)}$. A Harris ergodic Markov chain is uniformly ergodic if and only if it is uniformly mixing; see pp 367–368 in [28].

As with asymptotically uncorrelated sequences it is a consequence of the strong Markov property that if a Harris ergodic Markov chain is ϕ-mixing then it enjoys exponentially fast ϕ-mixing [4, Theorem 4.2] in the sense that there exists a $\theta > 0$ such that $\phi(n) = O(e^{-\theta n})$.

We collect and concisely state the main conclusions of this section.

Theorem 2. *Let X be a Harris ergodic Markov chain with stationary distribution π.*

1. *X is strongly mixing, i.e., $\alpha(n) \to 0$.*

2. *If* (3) *holds with* $E_\pi M < \infty$ *then* $\alpha(n) = O(\gamma(n))$.
3. *If* X *is geometrically ergodic and* (8) *holds then* X *is asymptotically uncorrelated, in which case there exists a* $\theta > 0$ *such that* $\rho(n) = O(e^{-\theta n})$.
4. X *is uniformly ergodic if and only if* X *is uniformly mixing, in which case there exists a* $\theta > 0$ *such that* $\phi(n) = O(e^{-\theta n})$.

4. Central Limit Theorems

We begin with a characterization of the CLT for strongly mixing processes. Define $S_n = \sum_{i=1}^n Y_i$ and $\sigma_n^2 = ES_n^2$.

Theorem 3. *[9, 14, 45] Let* Y *be a centered strictly stationary strongly mixing sequence such that* $EY_0^2 < \infty$. *If* $\sigma_n^2 \to \infty$ *as* $n \to \infty$ *then the following are equivalent:*

1. $S_n/\sigma_n \xrightarrow{d} N(0,1)$.
2. $\{S_n^2/\sigma_n^2\,,\ n \geq 1\}$ *is uniformly integrable.*

Remark 5. Since Harris ergodic Markov chains are strongly mixing this result is applicable in MCMC settings.

Remark 6. The assumption of stationarity is not an issue for Harris ergodic Markov chains since if a CLT holds for any one initial distribution then it holds for every initial distribution [41, Proposition 17.1.6].

Theorem 4. *[7] Let* X *be a Harris ergodic Markov chain and* f *be a function such that* $E_\pi f = 0$ *and* $E_\pi f^2 < \infty$. *Then the following are equivalent:*

1. $\sqrt{n}\bar{f}_n \xrightarrow{d} N(0,\sigma^2)$ *for some* $\sigma^2 \geq 0$.
2. $\{\sqrt{n}\bar{f}_n\,,\ n \geq 1\}$ *is bounded in probability.*

Remark 7. Another equivalence is given in [7] in terms of quantities based on the so-called *split chain*. But this is not germane to the current discussion.

4.1. Sufficient Conditions

Theorem 5. *[27, 28] Let* Y *be a centered strictly stationary strongly mixing sequence. Suppose at least one of the following conditions:*

1. *There exists* $B < \infty$ *such that* $|Y_n| < B$ *a.s. and* $\sum_n \alpha(n) < \infty$; *or*
2. $E|Y_n|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and*

$$\sum_n \alpha(n)^{\delta/(2+\delta)} < \infty\ . \tag{10}$$

Then

$$\sigma^2 = E(Y_0^2) + 2\sum_{j=1}^{\infty} E(Y_0 Y_j) < \infty$$

and if $\sigma^2 > 0$, *as* $n \to \infty$,

$$n^{-1/2} S_n \xrightarrow{d} N(0, \sigma^2) \ .$$

Corollary 1. *Let* $f : \mathsf{X} \to \mathbb{R}$ *be a Borel function such that* $E_\pi |f(x)|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and suppose* X *is a Harris ergodic Markov chain with stationary distribution* π. *If* (3) *holds such that* $E_\pi M < \infty$ *and* $\gamma(n)$ *satisfies*

$$\sum_n \gamma(n)^{\delta/(2+\delta)} < \infty \tag{11}$$

then for any initial distribution, as $n \to \infty$

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) \ .$$

Later, CLTs for ϕ-mixing and ρ-mixing Markov chains will be presented. However, the proofs of these results are similar to the proof of Corollary 1. Hence only the following proof is included.

Proof. Let $\alpha(n)$ and $\alpha_f(n)$ denote the strong mixing coefficients for the Markov chain $X = \{X_n\}$ and the functional process $\{f(X_n)\}$, respectively. By an earlier remark $\alpha_f(n) \le \alpha(n)$ for all $n \ge 1$. Moreover, we have that $\alpha(n) \le \gamma(n) \mathrm{E}_\pi M$ where $\gamma(n)$ and M are given in (3). Hence (11) guarantees that

$$\sum_n \alpha_f(n)^{\delta/(2+\delta)} < \infty$$

and the result follows from the Theorem and Remark 6. $\square$

Corollary 1 immediately yields some special cases which have proven to be useful in MCMC settings.

Corollary 2. *Suppose* X *is a Harris ergodic Markov chain with stationary distribution* π *and let* $f : \mathsf{X} \to \mathbb{R}$ *be a Borel function. Assume one of the following conditions:*

1. *[6]* X *is geometrically ergodic and* $E_\pi |f(x)|^{2+\delta} < \infty$ *for some* $\delta > 0$;
2. X *is polynomially ergodic of order* m, $E_\pi M < \infty$ *and* $E_\pi |f(x)|^{2+\delta} < \infty$ *where* $m\delta > 2 + \delta$; *or*
3. X *is polynomially ergodic of order* $m > 1$, $E_\pi M < \infty$ *and there exists* $B < \infty$ *such that* $|f(x)| < B$ π*-almost surely.*

Then for any initial distribution, as $n \to \infty$

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) \ .$$

For geometrically ergodic Markov chains the moment condition can not be weakened to a second moment (i.e., $\mathrm{E}_\pi f^2(x) < \infty$) without additional assumptions. See [24] for a construction that establishes the existence of a geometrically ergodic Markov chain and a function f such that $\mathrm{E}_\pi f^2(x) < \infty$ yet a CLT fails for any choice of σ^2. Also, see [2] for a counterexample with the same conclusion. These results are not too surprising since there are non-trivial counterexamples that indicate that the conditions of Theorem 5 are nearly as good as can be expected. For example, in [25] there is a construction of a strictly stationary sequence of uncorrelated random variables, $\{Y_n\}$, that have an arbitrarily fast strong mixing rate and $0 < EY_1^2 < \infty$ yet the CLT fails. Further counterexamples have been given in [12] and [3]. However, a slightly weaker moment condition is available if the sequence enjoys at least exponentially fast strong mixing which is the case for geometrically ergodic Markov chains. The following theorem is a special case of a result in [17].

Theorem 6. *[17] Let Y be a centered strictly stationary strongly mixing sequence. If the strong mixing coefficients satisfy $\alpha(n) = O(a^n)$ for some $0 < a < 1$ and $E[Y_0^2(\log^+ |Y_0|) < \infty$ then*

$$\sigma^2 = EY_0^2 + 2\sum_{k=1}^{\infty} E(Y_0 Y_k)$$

converges absolutely and if $\sigma^2 > 0$, as $n \to \infty$

$$n^{-1/2} S_n \xrightarrow{d} N(0, \sigma^2) .$$

Corollary 3. *Suppose X is a Harris ergodic Markov chain with stationary distribution π and let $f : \mathsf{X} \to \mathbb{R}$ be a Borel function. If X is geometrically ergodic and $E_\pi[f^2(x)(\log^+ |f(x)|)] < \infty$ then for any initial distribution, as $n \to \infty$*

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) .$$

A weaker moment condition is available for ρ–mixing sequences.

Theorem 7. *[29] Let Y be a centered strictly stationary ρ–mixing sequence with $EY_0^2 < \infty$. Suppose*

$$\sum_{n=1}^{\infty} \rho(n) < \infty . \tag{12}$$

Then

$$\sigma^2 = EY_0^2 + 2\sum_{k=1}^{\infty} E(Y_0 Y_k)$$

converges absolutely and if $\sigma^2 > 0$, as $n \to \infty$

$$n^{-1/2} S_n \xrightarrow{d} N(0, \sigma^2) .$$

Recall that if the Markov chain X is geometrically ergodic and satisfies detailed balance, it enjoys exponentially fast ρ–mixing and hence (12) obtains.

Corollary 4. *[54] Let X be a geometrically ergodic Markov chain with stationary distribution π. Suppose X satisfies* (8) *and that $E_\pi f^2(x) < \infty$. Then for any initial distribution, as $n \to \infty$*

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) .$$

Remark 8. In [54] this result is obtained via Corollary 1.5 in [35]. We have thus provided an alternative derivation.

An accessible proof of the following result may be found in [1] and [28]. Also see Chapter 5 of [15] and Lemma 3.3 in [10].

Theorem 8. *Let Y be a centered strictly stationary uniformly mixing sequence with $EY_0^2 < \infty$. If*

$$\sum_n \sqrt{\phi(n)} < \infty \tag{13}$$

then

$$\sigma^2 = EY_0^2 + 2\sum_{k=1}^{\infty} E(Y_0 Y_k)$$

converges absolutely and if $\sigma^2 > 0$ then as $n \to \infty$

$$n^{-1/2} S_n \xrightarrow{d} N(0, \sigma^2) .$$

If X is uniformly ergodic the coefficients $\phi(n)$ decrease exponentially and (13) is obvious.

Corollary 5. *[28, 62] Let X be a uniformly ergodic Markov chain with stationary distribution π. Suppose $E_\pi f^2(x) < \infty$. Then for any initial distribution, as $n \to \infty$*

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) .$$

The main conclusions of this section can be concisely stated as follows.

Theorem 9. *Let X be a Harris ergodic Markov chain on X with invariant distribution π and let $f : \mathsf{X} \to \mathbb{R}$ is a Borel function. Assume one of the following conditions:*

1. *X is polynomially ergodic of order $m > 1$, $E_\pi M < \infty$ and there exists $B < \infty$ such that $|f(x)| < B$ almost surely;*
2. *X is polynomially ergodic of order m, $E_\pi M < \infty$ and $E_\pi |f(x)|^{2+\delta} < \infty$ where $m\delta > 2 + \delta$;*
3. *X is geometrically ergodic and $E_\pi |f(x)|^{2+\delta} < \infty$ for some $\delta > 0$;*
4. *X is geometrically ergodic and $E_\pi[f^2(x)(\log^+ |f(x)|)] < \infty$;*
5. *X is geometrically ergodic, satisfies detailed balance and $E_\pi f^2(x) < \infty$; or*
6. *X is uniformly ergodic and $E_\pi f^2(x) < \infty$.*

Then for any initial distribution, as $n \to \infty$

$$\sqrt{n}(\bar{f}_n - E_\pi f) \xrightarrow{d} N(0, \sigma_f^2) .$$

Remark 9. Condition 1 of the theorem is interesting for applications of MCMC in Bayesian settings. In this case, it is often the case that posterior probabilities, i.e. expectations of indicator functions, are of interest. Since indicator functions are bounded it follows that a CLT will hold under a very weak mixing condition.

5. Examples

5.1. Toy Examples Revisited

Example 1 continued. Recall that since X is finite this chain is uniformly ergodic and uniformly mixing. Hence Corollary 5 implies that a CLT will hold for $\bar{f}_n$ if $\mathrm{E}_\pi f^2(x) < \infty$ which will hold except in unusual cases.

Example 2 continued. This chain is geometrically ergodic but does not satisfy (8). Hence it is strongly mixing and we can not conclude that it is asymptotically uncorrelated. Thus the best we can do is to appeal to Corollary 3 and conclude that a CLT will hold for $\bar{f}_n$ if $\mathrm{E}_\pi[f(x)^2(\log^+|f(x)|)] < \infty$. Recall that in subsection 2.1 it was shown that a CLT holds for $\bar{f}_n$ if $f^2(x) \le a^{|x|}\ \forall\, x \in \mathbb{Z}$ when $a > 1$ satisfies $a\theta < 1$ and $(a\theta - 1)a + 1 - \theta < 0$.

Example 3 continued. Let $m > 2$ and recall that this random walk is polynomially ergodic of order $m - 1$ and that Theorem 1 says a CLT holds if $f(x) \le (x+1)^{m(\tau+\eta-1)}$ for all $x \ge 0$ where $1 - \tau \le \eta \le 1$ is such that $\mathrm{E}_\pi(x+1)^{2m\eta} < \infty$. Alternatively, an appeal to Corollary 2 says that we have a CLT if $\mathrm{E}_\pi(x+1)^m < \infty$ and $\mathrm{E}_\pi|f(x)|^{2+\delta} < \infty$ where $\delta > 2/(m-2)$.

5.2. A Benchmark Gibbs Sampler

The following Gibbs sampler is similar to one used by many authors to analyze the benchmark pump failure data set in [20]. For example, in [51], [60], and [62] similar settings are considered and uniform ergodicity of the corresponding Gibbs samplers is established.

Set $y = (y_1, y_2, \ldots, y_n)^T$ and let $\pi(x, y)$ be a joint density on $\mathbb{R}^{n+1}$ such that the corresponding full conditionals are

$$X|y \sim \mathrm{Gamma}\,(\alpha_1, a + b^T y)$$
$$Y_i|x \sim \mathrm{Gamma}\,(\alpha_{2i}, \beta_i(x))$$

for $i = 1, \ldots, n$, $b = (b_1, \ldots, b_n)^T$ where $a > 0$ and each $b_i > 0$ are known. (Say $U \sim \mathrm{Gamma}(\alpha, \beta)$ if its density is proportional to $u^{\alpha-1}e^{-u\beta}I(u > 0)$.) Since, conditional on x, the order in which the Y_i are updated is irrelevant we will use a two variable Gibbs sampler with the transition rule $(x', y') \to (x, y)$; that is, given that the current value is $(X_n = x', Y_n = y')$ we obtain (X_{n+1}, Y_{n+1}) by first drawing $x \sim f(X_{n+1}|y')$ then $Y_{i,n+1} \sim f(y_i|x)$. A minor modification of the argument in [62] will show that (4) holds on $C = \mathsf{X}$ with $n_0 = 1$ and if for

$i = 1, \dots, n$ there is a function $g > 0$ such that for all $x > 0$

$$\frac{\beta_i(x)}{b_i x + \beta_i(x)} \geq g(x) \; . \tag{14}$$

Thus if (14) holds this Gibbs sampler is uniformly ergodic (or uniformly mixing) and an appeal to Corollary 5 shows that a CLT is assured if $\mathrm{E}_\pi f^2(x) < \infty$.

5.3. A Gibbs Sampler for a Hierarchical Bayes Setting

Consider the following Bayesian version of the classical normal theory one–way random effects model. First, conditional on $\theta = (\theta_1, \dots, \theta_K)^T$ and λ_e, the data, Y_{ij}, are independent with

$$Y_{ij} | \theta, \lambda_e \sim \mathrm{N}(\theta_i, \lambda_e^{-1})$$

where $i = 1, \dots, K$ and $j = 1, \dots, m_i$. Conditional on μ and λ_θ, $\theta_1, \dots, \theta_K$ and λ_e are independent with

$$\theta_i | \mu, \lambda_\theta \sim \mathrm{N}(\mu, \lambda_\theta^{-1}) \quad \text{and} \quad \lambda_e \sim \mathrm{Gamma}(a_2, b_2),$$

where a_2 and b_2 are known positive constants. Finally, μ and λ_θ are assumed independent with

$$\mu \sim \mathrm{N}(m_0, s_0^{-1}) \quad \text{and} \quad \lambda_\theta \sim \mathrm{Gamma}(a_1, b_1)$$

where m_0, s_0, a_1 and b_1 are known constants; all of the priors are proper since s_0, a_1 and b_1 are assumed to be positive and $m_0 \in \mathbb{R}$. The posterior density of this hierarchical model is characterized by

$$\pi(\theta, \mu, \lambda | y) \propto g(y|\theta, \lambda_e) g(\theta|\mu, \lambda_\theta) g(\lambda_e) g(\mu) g(\lambda_\theta) \tag{15}$$

where $\lambda = (\lambda_\theta, \lambda_e)^T$, y is a vector containing all of the data, and g denotes a generic density. Expectations with respect to π typically require evaluation of ratios of intractable integrals, which may have dimension $K + 3$ and typically, $K \geq 3$.

We are interested in the standard Gibbs sampler which leaves the posterior (15) invariant. Define

$$v_1(\theta, \mu) = \sum_{i=1}^{K} (\theta_i - \mu)^2 \; , \qquad v_2(\theta) = \sum_{i=1}^{K} m_i (\theta_i - \overline{y}_i)^2 \quad \text{and} \quad \mathrm{SSE} = \sum_{i,j} (y_{ij} - \overline{y}_i)^2$$

where $\overline{y}_i = m_i^{-1} \sum_{j=1}^{m_i} y_{ij}$. The full conditionals for the variance components are

$$\lambda_\theta | \theta, \mu, \lambda_e, y \sim \mathrm{Gamma}\left(\frac{K}{2} + a_1, \frac{v_1(\theta, \mu)}{2} + b_1 \right) \tag{16}$$

and

$$\lambda_e | \theta, \mu, \lambda_\theta, y \sim \mathrm{Gamma}\left(\frac{M}{2} + a_2, \frac{v_2(\theta) + \mathrm{SSE}}{2} + b_2 \right) \tag{17}$$

where $M = \sum_i m_i$. If $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_K)^T$ and $\overline{\theta} = K^{-1} \sum_i \theta_i$, the remaining full conditionals are

$$\theta_i | \theta_{-i}, \mu, \lambda_\theta, \lambda_e, y \sim \; \mathrm{N}\left(\frac{\lambda_\theta \mu + m_i \lambda_e \overline{y} i}{\lambda_\theta + m_i \lambda_e}, \frac{1}{\lambda_\theta + m_i \lambda_e} \right)$$

for $i = 1, \ldots, K$ and

$$\mu | \theta, \lambda_\theta, \lambda_e, y \sim \; \mathrm{N}\left(\frac{s_0 m_0 + K \lambda_\theta \overline{\theta}}{s_0 + K \lambda_\theta}, \frac{1}{s_0 + K \lambda_\theta} \right) .$$

Our fixed-scan Gibbs sampler updates μ then the θ_i's then λ_θ and λ_e. Since the θ_i's are conditionally independent given (μ, λ), the order in which they are updated is irrelevant. The same is true of λ_θ and λ_e since these two random variables are conditionally independent given (θ, μ). A one–step transition of this Gibbs sampler is $(\mu', \theta', \lambda') \to (\mu, \theta, \lambda)$ meaning that we sequentially draw from the conditional distributions $\mu | \lambda', \theta'$ then $\theta_i | \theta_{-i}, \mu, \lambda'$ for $i = 1, \ldots, K$ then $\lambda_e | \mu, \theta$ then $\lambda_\theta | \mu, \theta$. Assume that $m' = \min\{m_1, m_2, \ldots, m_K\} \geq 2$ and that $K \geq 3$. Let $m'' = \max\{m_1, m_2, \ldots, m_K\}$. Define $\delta_1 = (2a_1 + K - 2)^{-1}$ and $\delta_2 = (2a_1 - 2)^{-1}$.

Proposition 1. *[34] Assume that $a_1 > 3/2$ and $5m' > m''$. Fix $c_1 \in \big(0, \min\{b_1, b_2\}\big)$. Then the Gibbs sampler satisfies* (5) *with the drift function*

$$V(\theta, \lambda) = 1 + e^{c_1 \lambda_\theta} + e^{c_1 \lambda_e} + \frac{\delta_2}{K \delta_1 \lambda_\theta} + \frac{K \lambda_\theta}{s_0 + K \lambda_\theta} (\overline{\theta} - \overline{y})^2 .$$

Remark 10. Values for the constants in (5) are given in [34] but in an effort to keep the notation under control we do not report them here.

Theorem 1 immediately implies a CLT for $\bar{f}_n$ for any function f such that $f^2(\mu, \theta, \lambda) \leq V(\theta, \lambda)$ for all (μ, θ, λ). Of course, it is easy to find functions involving μ or θ that do not satisfy this requirement. On the other hand, Theorem 1 will be useful for many functions of only λ_θ and λ_e.

Recall that the drift function may not be unique. Prior to the work of [34], this Gibbs sampler was also analyzed in [26] wherein (5) was established using a different drift function and more restrictive conditions on a_1 and m'. However, this drift function can alleviate some of the difficulties with using Theorem 1 for functions involving μ.

Proposition 2. *[26] Assume that $a_1 \geq (3K-2)/(2K-2)$ and $m' > (\sqrt{5}-2)m''$. Then the Gibbs sampler satisfies* (5) *with the drift function*

$$\begin{aligned} W(\mu, \theta, \lambda) =& 1 + \vartheta \left(\frac{1}{\lambda_e} + \sum_{i=1}^{K} m_i (\bar{y}_i - \theta_i)^2 + (\mu - \bar{y})^2 \right) + \\ &+ \frac{1}{\lambda_\theta} + e^{c_1 \lambda_\theta} + e^{c_1 \lambda_e} + \frac{K \lambda_\theta}{s_0 + K \lambda_\theta} (\overline{\theta} - \overline{y})^2 . \end{aligned}$$

where $0 < \vartheta < 1$ is a constant defined on p. 427 in [26].

Proposition 1 shows that this Gibbs sampler is geometrically ergodic as long as $a_1 > 3/2$ and $5m' > m''$. However, it does not satisfy detailed balance. An appeal to Corollary 2 or 3 shows that functions with a little bit more than a second moment with respect to (15) will enjoy a CLT.

5.4. Independence Sampler

The independence sampler is an important special case of the MHG algorithm. Suppose the target distribution π has support $\mathsf{X} \subseteq \mathbb{R}^k$ and a density which, in a slight abuse of notation, is also denoted π. Let p be a proposal density whose support contains X and suppose the current state of the chain is $X_n = x$. Draw $y \sim p$ and set $X_{n+1} = y$ with probability

$$\alpha(x, y) = \frac{\pi(y)p(x)}{\pi(x)p(y)} \wedge 1 \; ;$$

otherwise set $X_{n+1} = x$. This Markov chain is Harris ergodic and it is well-known [40] that it is uniformly ergodic if there exists $\kappa > 0$ such that

$$\frac{\pi(x)}{p(x)} \leq \kappa \tag{18}$$

for all x since (18) implies a minorization (4) on X with $n_0 = 1$ and $\epsilon = 1/\kappa$. Hence Corollary 5 implies that a CLT will hold for $\bar{f}_n$ if $\mathrm{E}_\pi f^2 < \infty$. On the other hand, the independence sampler will not even be geometrically ergodic if there is a set of positive π-measure where (18) fails to hold. Moreover, in this case there are conditions given in [52] which ensure a CLT *cannot* hold.

5.5. An MHG Algorithm for Finite Point Processes

The material in this subsection is adapted from [22] and [44]. Let $\mathcal{X}$ be a bounded region of $\mathbb{R}^d$ and let λ be Lebesgue measure. Define $\mathcal{X}^0 := \{\varnothing\}$ and for $k \geq 1$, $\mathcal{X}^k := \mathcal{X} \times \cdots \times \mathcal{X}$ (there being k terms in the Cartesian product). Think of $x \in \mathcal{X}^k$ as a pattern of k points in $\mathcal{X}$, in particular, $\mathcal{X}^0$ denotes the pattern with no points, and define $n(x)$ to be the cardinality of x so that if $x \in \mathcal{X}^k$ then $n(x) = k$. Let the state space X be the union of all $\mathcal{X}^k$, that is, $\mathsf{X} = \cup_{i=0}^{\infty} \mathsf{X}_i$ where $\mathsf{X}_i = \{x : n(x) = i\}$. The target π is an unnormalized density with respect to the Poisson process with intensity measure λ on $\mathcal{X}$. The following MHG algorithm for simulating from π is proposed in [22]:

1. With probability 1/2 attempt an up step
   (a) Draw $\xi \sim \lambda(\cdot)/\lambda(\mathcal{X})$. Set $x = x \cup \xi$ with probability

$$1 \wedge \frac{\lambda(\mathcal{X})\,\pi(x \cup \xi)}{(n(x)+1)\,\pi(x)} \; .$$

2. Else attempt a down step
   (a) If $x = \varnothing$ skip the down step
   (b) Draw ξ uniformly from the points of x. Set $x = x \setminus \xi$ with probability

$$1 \wedge \frac{n(x)\,\pi(x \setminus \xi)}{\lambda(\mathcal{X})\,\pi(x)} \, .$$

This MHG algorithm is Harris recurrent and geometrically ergodic.

Proposition 3. *[22] Suppose there exists a real number M such that*

$$\pi(x \cup \xi) \leq M\pi(x)$$

for all $x \in \mathsf{X}$ and all $\xi \in \mathcal{X}$. Then the MHG algorithm started at $x^* \in \{x : \pi(x) > 0\}$ is Harris ergodic and satisfies (5) *with the drift function $V(x) = A^{n(x)}$ where $A > M\lambda(\mathcal{X}) \vee 1$.*

Of course, Theorem 1 implies a CLT for $\bar{f}_n$ for any function f such that $f^2(x) \leq A^{n(x)}$ for all x. On the other hand, this algorithm was constructed so as to satisfy (8) (see [22] for a detailed argument) and hence the Markov chain is asymptotically uncorrelated so that a CLT holds when $\mathrm{E}_\pi f^2(x) < \infty$.

5.6. Random Walk MHG Algorithms

Let π be a target density on $\mathbb{R}^k$ and let the proposal density have the form $q(y|x) = q(|y-x|)$. Now suppose that the current state of the chain is $X_n = x$. Draw $y \sim q$ and set $X_{n+1} = y$ with probability

$$\alpha(x,y) = \frac{\pi(y)}{\pi(x)} \wedge 1 \; ;$$

otherwise set $X_{n+1} = x$. Note that this algorithm satisfies (8) by construction.

Random walk-type MHG algorithms are some of the most useful and popular MCMC algorithms and consequently their theoretical properties have been thoroughly studied. In [40] it was shown that random walk samplers (on $\mathbb{R}^k$) *cannot* be uniformly ergodic (or uniformly mixing) but they do establish that a random walk MHG algorithm can be geometrically ergodic by verifying (5) when $k = 1$ and π has tails that decrease exponentially. This work was extended in [58] by establishing (5) in the case where $k \geq 1$. However, in [30] the drift (5) was verified with a different drift function than that used in [58] and consequently obtained more general conditions ensuring geometric ergodicity. On the other hand, if a random walk MHG algorithm is not geometrically ergodic it may still be polynomially ergodic of all orders; see [18].

Proposition 4. *[30] Suppose π is a positive density on $\mathbb{R}^k$ having continuous first derivatives such that*

$$\lim_{|x| \to \infty} \frac{x}{|x|} \cdot \nabla \log \pi(x) = -\infty \, .$$

Let $A(x) := \{y \in \mathbb{R}^k : \pi(y) \geq \pi(x)\}$ be the region of certain acceptance and assume that there exist $\delta > 0$ and $\epsilon > 0$ such that, for every x, $|x - y| \leq \delta$ implies $q(y|x) \geq \epsilon$. Then if

$$\liminf_{|x| \to \infty} \int_{A(x)} q(y|x)\, dy > 0$$

the random walk MHG algorithm satisfies (5) *with the drift function $V(x) = c\pi(x)^{-1/2}$ for some $c > 0$.*

Hence, under the conditions of Proposition 4, Theorem 1 guarantees a CLT if $f(x)$ satisfies $f^2(x) \leq c\pi(x)^{-1/2}$ for all $x \in \mathbb{R}^k$. Alternatively, we conclude that the random walk MHG is geometrically ergodic, satisfies (8) (and hence is asymptotically uncorrelated) and an appeal to Corollary 4 establishes the existence of a CLT if $\mathrm{E}_\pi f^2(x) < \infty$.

6. Final Remarks

The focus has been on some of the connections between recent work on general state space Markov chains and results from mixing processes and the implications for Markov chain CLTs. However, this article only scratches the surface of the mixing process literature that is potentially useful in MCMC. For example, the existence of a functional CLT or strong invariance principle is required in order to estimate σ_f^2 from (1) [11, 23, 32]. There has been much work on these for mixing processes; a good starting place for strong invariance principles is [49] while an introduction to the functional CLT is contained in [1].

Appendix A: Calculations for Example 2

Define $V(z) = a^{|z|}$ for some $a > 1$. Then $V(z) \geq 1$ for all $z \in \mathbb{Z}$ and

$$PV(x) = \sum_{y \in \mathbb{Z}} a^{|y|} P(x, y) \ .$$

Recall that $\Delta V(x) = PV(x) - V(x)$. The first goal is to show that if $x \neq 0$ then $\Delta V(x)/V(x) < 0$ since then there must be a $\beta > 0$ such that $\Delta V(x) < -\beta V(x)$. Suppose $X_n = x \geq 1$ then

$$PV(x) = \theta a^{x+1} + 1 - \theta \ \Rightarrow \ \frac{\Delta V(x)}{V(x)} = a\theta - 1 + \frac{1-\theta}{a^x} < 0$$

as long as

$$(a\theta - 1)V(x) + 1 - \theta < 0 \ . \tag{19}$$

Now (19) can hold only if $a\theta - 1 < 0$ and since $V(x) \geq a$ for all $x \neq 0$ (19) will hold when $a\theta - 1 < 0$ and $(a\theta - 1)a + 1 - \theta < 0$. A similar argument shows that this is also the case when $X_n = -x \leq -1$. Now suppose $X_n = 0$.

Then $PV(0) = a$ and $\Delta V(0) = -V(0) + a$. Putting this together yields (5) with $C = \{0\}$.

Acknowledgments

The author is grateful to Charlie Geyer, Murali Haran, Jeff Rosenthal and Qi-Man Shao for helpful conversations about this paper. Also, the author thanks an anonymous referee for many very helpful comments.

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York. MR233396

[2] Bradley, R. C. (1983). Information inequality and the central limit question. *Rocky Mountain Journal of Mathematics* 13: 77-97. MR692579

[3] Bradley, R. C. (1985). On the central limit question under absolute regularity. *The Annals of Probability*, 13:1314–1325. MR806228

[4] Bradley, R. C. (1986). Basic properties of strong mixing conditions. In Eberlein, E. and Taqqu, M. S., editors, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 165–192. Birkhauser, Cambridge, MA. MR899990

[5] Bradley, R. C. (1999). Can a theorem of Csáki and Fischer provide a key to Ibragimov's conjecture? In *Bolyai Society Mathematical Studies 9, Random Walks, Budapest (Hungary) 1999*, pages 11–42. North-Holland, Amsterdam. MR1752729

[6] Chan, K. S. and Geyer, C. J. (1994). Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758. MR1329179

[7] Chen, X. (1999). Limit theorems for functionals of ergodic Markov chains with general state space. *Memoirs of the American Mathematical Society*, 139. MR1491814

[8] Christensen, O. F., Moller, J., and Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodology and Computing in Applied Probability*, 3:309–327. MR1891114

[9] Cogburn, R. (1960). Asymptotic properties of stationary sequences. *University of California Publications in Statistics*, 30, 63–82. MR138118

[10] Cogburn, R. (1972). The central limit theorem for Markov processes. In Le Cam, L. E., Neyman, J., and Scott, E. L., editors, *Proceedings of the Sixth Annual Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 485–512. University of California Press, Berkeley. MR402869

[11] Damerdji, H. (1994). Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19:494–512. MR1290511

[12] Davydov, Y. A. (1973). Mixing conditions for Markov chains. *Theory of Probability and Its Applications*, 27:312–328.

[13] Dehling, H., Denker, M., and Philipp, W. (1986). Central limit theorems for mixing sequences of random variables under minimal conditions. *The Annals of Probability*, 14:1359–1370. MR866356

[14] Denker, M. (1986). Uniform integrability and the central limit theorem. In Eberlein, E. and Taqqu, M. S., editors, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 269–274. Birkhauser, Cambridge, MA. MR899993

[15] Doob, J. L. (1953). *Stochastic Processes*. John Wiley & Sons, New York. MR58896

[16] Douc, R., Fort, G., Moulines, E., and Soulier, P. (2204). Practical drift conditions for subgeometric rates of convergence. *The Annals of Applied Probability*, 14:1353–1377. MR2071426

[17] Doukhan, P., Massart, P., and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Annals of the Institute Henri Poincaré, Probability and Statistics*, 30: 63-82. MR1262892

[18] Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters*, 49:401–410. MR1796485

[19] Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and their Applications*, 103:57–99. MR1947960

[20] Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analyses of event rates. *Technometrics*, 29:1–15. MR876882

[21] Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, 7:473–511.

[22] Geyer, C. J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, pages 79–140. Chapman & Hall/CRC, Boca Raton. MR1673118

[23] Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2:180–198. MR1143399

[24] Häggström, O. (2004). On the central limit theorem for geometrically ergodic Markov chains. *Probability Theory and Related Fields* (to appear). MR2027296

[25] Herndorf, N. (1983). Stationary strongly mixing sequences not satisfying the central limit theorem. *The Annals of Probability*, 11:809–813. MR704571

[26] Hobert, J. P. and Geyer, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67:414–430. MR1659196

[27] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and Its Applications*, 7:349–382. MR148125

[28] Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff, The Netherlands. MR322926

[29] Ibragimov, I. A., (1975) A note on the central limit theorem for dependent random variables. *Theory of Probability and its Applications* 20: 135-141. MR362448

[30] Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications*, 85:341–361. MR1731030

[31] Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *The Annals of Applied Probability*, 12:224–247. MR1890063

[32] Jones, G. L., Haran, M., and Caffo, B. S. (2004). Output analysis for Markov chain Monte Carlo simulations. Technical report, University of Minnesota, School of Statistics. MR2043391

[33] Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334. MR1888447

[34] Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32:784–817. MR2060178

[35] Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104:1–19. MR834478

[36] Kontoyiannis, I. and Meyn, S. P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *The Annals of Applied Probability*, 13:304–362. MR1952001

[37] Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley, New York. MR1180522

[38] Liu, J. S., Wong, W. H., and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society,* Series B, 57:157–169. MR1325382

[39] Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *Journal of the American Statistical Association*, 99:228–238. MR2054301

[40] Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121. MR1389882

[41] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London. MR1287609

[42] Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, 4:981–1011. MR1304770

[43] Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, 29:1–12. MR1894377

[44] Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, pages 141–172. Chapman & Hall/CRC, Boca Raton. MR1673122

[45] Mori, T. and Yoshihara, K. (1986). A note on the central limit theorem for stationary strong mixing sequences. *Yokohama Mathematical Journal*, 34: 143–146. MR886062

[46] Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, London. MR776608

[47] Nummelin, E. (2002). MC's for MCMC'ists. *International Statistical Review*, 70:215–240.

[48] Peligrad, M. (1982). Invariance principles for mixing sequences of random variables. *The Annals of Probability*, 10:968–981. MR672297

[49] Philipp, W. and Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoirs of the American Mathematical Society*, 2:1–140. MR433597

[50] Robert, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science*, 10:231–253. MR1390517

[51] Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York. MR1707311

[52] Roberts, G. O. (1999). A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. *Journal of Applied Probability*, 36:1210–1217. MR1742161

[53] Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 56:377–384. MR1281941

[54] Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25. MR1448322

[55] Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society,* Series B, 61:643–660. MR1707866

[56] Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, 28:489–504. MR1858413

[57] Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71.

[58] Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110. MR1399158

[59] Rosenblatt, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. Springer-Verlag, New York. MR329037

[60] Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566. MR1340509

[61] Rosenthal, J. S. (1996). Analysis of the Gibbs sampler for a model related

to James-Stein estimators. *Statistics and Computing*, 6:269–275.

[62] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762. MR1329166

[63] Tuomimem, P. and Tweedie, R. L. (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Advances in Applied Probability*, 26:775–798. MR1285459