Variable-at-a-time Markov chain Monte Carlo

Galin Jones[1]

School of Statistics
University of Minnesota

August 2010

[1] joint work with Charles Geyer, Alicia Johnson and Ron Neath

General Setting

Let ϖ be a probability distribution with support contained in $\mathbb{R}^d$.

We want to calculate features of ϖ such as $E_\varpi h(X)$.

Basic idea behind MCMC: Construct a Markov chain

$$X = \{X_0, X_1, X_2, \ldots\}$$

having ϖ as its invariant distribution—if

$$P^n(x, A) := \Pr(X_{i+n} \in A | X_i = x) \; ,$$

then $\varpi P = \varpi$. Frequently,

$$P(x', A) = \int_A k(x|x')\varphi(dx)$$

Useful MCMC

Minimal requirements for useful MCMC: As $n \to \infty$

- $X_n \xrightarrow{d} X_\varpi$ and
- a LLN holds
$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to E_\varpi h(X) \; .$$

Useful MCMC

Assume $X = \{X_n\}$ is

- ϖ-invariant,
- φ-irreducible.

If $E_\varpi |h(X)| < \infty$, then as $n \to \infty$

$$\bar{h}_n = \frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \stackrel{a.s}{\to} E_\varpi h(X)\,,$$

If, in addition, X is aperiodic, then for φ-almost all x, as $n \to \infty$

$$\|P^n(x_0, \cdot) - \varpi(\cdot)\| := \sup_A |P^n(x_0, A) - \varpi(A)| \downarrow 0$$

Assumptions

φ-irreducible :

There exists a σ-finite measure φ such that

$$\varphi(A) > 0 \Rightarrow P^n(x, A) > 0 \quad \text{for some } n = n(x, A) \, .$$

aperiodic:

There do not exist disjoint subsets $A_1, A_2, \ldots, A_d$, N such that

$$P(x, A_{i+1}) = 1 \quad \text{for all } x \in A_i \, .$$

φ-irreducibility appears easier to verify and several authors have observed that periodicity appears to be “rather rare” in MCMC.

To what extent does φ-irreducibility imply aperiodicity?

Toy Example

Suppose ϖ is the uniform distribution on two points $\{0, 1\}$.

Let the Markov kernel be

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

It is easy to check that

$$\varpi P = (.5, .5) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (.5, .5) = \varpi$$

and that P is irreducible but P is not aperiodic.

This is an example of a Metropolis algorithm where every proposal is accepted.

Basic Result

Standing Assumption: P is φ-irreducible.

Thm Suppose A is such that $\varphi(A) > 0$. Let $x \in A$ and assume that $P(x, B) > 0$ for every $B \subseteq A$ such that $x \in B$ and $\varphi(B) > 0$. Then P is aperiodic.

Cor If $P(x, \{x\}) > 0$ for some x, then P is aperiodic.

Thm: If there exists an x such that $P(x, \{x\}) > 0$, then MH is aperiodic.

The condition of the theorem is usually easy to verify but recall the two-point Metropolis example.

Variable-at-a-time MCMC

Suppose ϖ admits a density π of the form

$$\pi(x, y) = \pi(x_1, \ldots, x_{d_1}, y_1, \ldots, y_{d_2})$$

and $f(x|y)$ and $g(y|x)$ are the conditionals.

Gibbs Samplers Consider the one-step transition $(x', y') \to (x, y)$

DUGS: $f(x|y')g(y|x)$

RQGS: $af(x|y')g(y|x) + (1-a)g(y|x')f(y|x)$

RSGS: $af(x|y')\delta(y-y') + (1-a)g(y|x')\delta(x-x')$

Aperiodicity of Gibbs Samplers

The DUGS kernel is

$$P_{DUGS}((x', y'), A) = \int_A f(x|y')g(y|x)dxdy$$

Notice that $P_{DUGS}((x, y), \{x, y\}) = 0$

Does φ-irreducibility imply aperiodicity for Gibbs samplers?

Aperiodicity of Gibbs Samplers

Thm Suppose A is such that $\varphi(A) > 0$. Let $x \in A$ and assume that $P(x, B) > 0$ for every $B \subseteq A$ such that $x \in B$ and $\varphi(B) > 0$. Then P is aperiodic.

Notice that if there is one set A such that $\varphi(A) > 0$ and

$$f(x|y')g(y|x)$$

is continuous in (x, y) and positive on A, then DUGS is aperiodic.

It is also easy to show that a Gibbs sampler on a finite state space must be aperiodic since $P(x, \{x\}) > 0$ for every x.

Aperiodicity of Gibbs Samplers

We can do better.

Thm DUGS and RSGS are aperiodic.

Key idea—elementary Gibbs updates are idempotent: $P_i^2 = P_i$.

Other Variable-at-a-time Samplers

Thm Suppose P_1 is aperiodic and $P_2, \ldots, P_c$ are some other Markov kernels having invariant distribution ϖ. Then the mixture

$$P = \sum_{i=1}^{c} a_i P_i$$

is aperiodic where $a_1 > 0$ and $\sum_{i=1}^{c} a_i = 1$.

Example:
The random sequence Gibbs sampler RQGS is aperiodic—

$$P((x', y'), A) = \int_A [a_1 f(x|y') g(y|x) + (1 - a_1) g(y|x') f(x|y)] dx dy \ .$$

Useful MCMC

Rate of convergence: X is geometrically ergodic if

$$\|P^n(x_0, \cdot) - \varpi(\cdot)\| \le C(x_0)t^n \quad C(x_0) \ge 0 \text{ and } \; t \in (0,1)$$

A key sufficient condition for

- the existence of CLTs

$$\sqrt{n}(\bar{X}_n - E_\varpi X) \xrightarrow{d} \text{N}(0, \sigma^2)$$

- constructing consistent estimators of σ^2 and
- implementing rigorous (fixed-width) stopping rules.

Gibbs Samplers

DUGS: $f(x|y')g(y|x)$

RQGS: $af(x|y')g(y|x) + (1-a)g(y|x')f(y|x)$

RSGS: $af(x|y')\delta(y-y') + (1-a)g(y|x')\delta(x-x')$

THM If the DUGS is geometrically ergodic, then so are the random sequence Gibbs sampler and the random scan Gibbs sampler. Moreover, the drift rates are ordered

$$\lambda_{DUGS} < \lambda^2_{RQGS} < \lambda_{RQGS} < \lambda^2_{RSGS} < \lambda_{RSGS} .$$

Metropolis-Hastings within Gibbs

Let h_g be a MH sampler having invariant density $g(y|x)$.

MHWG: $f(x|y')h_g(y|x, y')$

RSMHWG: $af(x|y')\delta(y - y') + (1 - a)h_g(y|x', y')\delta(x - x')$

THM Suppose h_f corresponds to an independence sampler or a Metropolis random walk. If the MHWG is geometrically ergodic, then so is RSMHWG. Moreover, the drift rates are ordered

$$\lambda_{MHWG} < \lambda^2_{RSMHWG} < \lambda_{RSMHWG} \ .$$

Where are we?

- Useful MCMC algorithms produce a realization of a Markov chain which (minimally) satisfies two properties: For almost all starting values (1) $X_n \xrightarrow{d} X_{\varpi}$ and (2) a LLN holds.
- This requires verification of φ-irreducibility and aperiodicity of the Markov chain.
- For most MCMC samplers, we have shown that φ-irreducibility implies aperiodicity outside of trivial cases. This is not limited to the two-variable setting.
- Geometric ergodicity is a key sufficient condition for useful MCMC by allowing calculation of valid Monte Carlo standard errors.
- Developed conditions for geometric ergodicity of DUGS, RQGS, RSGS, MHWG and RSMHWG. This is limited to the two-variable setting.