Comm. Korean Math. Soc. **14** (1999), No. 3, pp. 627–633

FUNCTIONAL CENTRAL LIMIT THEOREMS FOR THE GIBBS SAMPLER

OESOOK LEE

ABSTRACT. Let the given distribution π have a log-concave density which is proportional to $\exp(-V(\mathbf{x}))$ on R^d. We consider a Markov chain induced by the method of Gibbs sampling having π as its invariant distribution and prove geometric ergodicity and the functional central limit theorem for the process.

1. Introduction

Let π be a probability distribution on R^d $(d \geq 2)$, and suppose that we are interested in estimating characteristics of it, such as $\pi(B)$ or $\int f d\pi$ for some measurable function f. Even when π is fully specified, one may have resort to methods like Markov chain Monte Carlo simulation, especially when it is not computationally tractable. In other words, one can construct an irreducible Markov chain having the distribution π as its invariant distribution and whose transition function is tractable. One of the widely applicable methods of constructing such Markov chains is the method of Gibbs sampling (see, e.g., Gelfand and Smith (1990), Geyer (1992), Hobert and Casella (1998), Roberts and Rosental (1998)).

Let π be the joint probability distribution of $\mathbf{X} = (X_1, X_2, \cdots, X_d)$. The transition from $\mathbf{x}_{n-1}$ to $\mathbf{x}_n$ is just to replace the jth component of $\mathbf{x}_{n-1}$ by drawing a realization from the distribution $p(x_j|\mathbf{X}(-j) = \mathbf{x}_{n-1}(-j))$, where $\mathbf{x}_{n-1}(-j)$ is equal to $\mathbf{x}_{n-1}$ with its jth component being omitted

Received February 28, 1999. Revised May 28, 1999.

1991 Mathematics Subject Classification: 60J05, 60F05.

Key words and phrases: Gibbs sampler, Markov chain, geometric ergodicity, functional central limit theorem.

This research is supported by the grants from STEPI made in the program year of 1998-99.

and $1 \leq j \leq d$, $j = n$ modulo d. That is, the updating of $\mathbf{x}_n$ is done componentwise and sequentially by drawing from the conditional distribution of the component given the remaining components being fixed. Markov chain $\mathbf{X}_n$ obtained by the method of Gibbs sampling is inhomogeneous but $\mathbf{X}_{nd}$ is a homogeneous Markov chain.

Asymptotic behavior of the induced Markov chain is crucial for the success of approximation. In practice, the speed of convergence to the invariant distribution is of interest and it is desirable that the Markov chain is geometrically ergodic. Central limit theorem gives the size of the Monte Carlo error. Proving geometric ergodicity is not only important in its own right but also the easiest and the most common method to establishing functional central limit theorem.

We note that Chan (1993), Tierney (1994), Athreya and et al. (1996), Hobert and Geyer (1998) and Hwang and Sheu (1998) have developed sufficient conditions for geometric ergodicity of certain Gibbs sampler. In Roberts and Tweedie (1996) and Mengersen and Tweedie (1996), Hastings and Metropolis algorithm is considered, where drift condition described by Meyn and Tweedie (1993) is used to show that Gibbs sampler is geometrically ergodic and to establish central limit theorems.

In this paper, we model the Gibbs sampler in terms of stochastic difference equation and use the drift condition to prove geometric ergodicity of the induced Markov chain, and then obtain a class of functions for which functional central limit theorem holds.

We refer the reader to Meyn and Tweedie (1993) for general contents of Markov chain theory and to Hwang and Sheu (1998) for stochastic difference equation approach for Gibbs sampler.

2. Main Results

Let π be a probability distribution on $(R^d, \mathcal{B}^d)$ known to a constant multiple and let π have a density given by

$$\pi(\mathbf{x}) \propto e^{-V(\mathbf{x})}. \tag{2.1}$$

Let $\mathbf{X}^{(0)}$ be a given initial point and let $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots$ be the Gibbs sampler, which is homogeneous Markov chain in R^d with transition probability density function $p(\mathbf{x}, \mathbf{y})$ given by, for $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ and

$\mathbf{y} = (y_1, y_2, \cdots, y_d)$,

$p(\mathbf{x}, \mathbf{y}) = p_1(y_1|(x_2, \cdots, x_d))p_2(y_2|(y_1, x_3, \cdots, x_d)) \cdots p_d(y_d|(y_1, \cdots, y_{d-1}))$

where

$$p_k(y_k|(y_1, \cdots, y_{k-1}, x_{k+1}, \cdots, x_d)) = \frac{e^{-V(y_1, \cdots, y_k, x_{k+1}, \cdots, x_d)}}{\int e^{-V(y_1, \cdots, y_k, x_{k+1}, \cdots, x_d)} dy_k}.$$

Let $P^{(n)}(\mathbf{x}, d\mathbf{y})$ denote the n-step transition probability function for $\{\mathbf{X}^{(n)}\}$.

We make the following assumptions on V:

Assumption (A)

(1) $V : R^d \to R$ is smooth, strictly convex and there are $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1 \leq \frac{\partial^2 V}{\partial x_i \partial x_j}(\mathbf{x}) \leq \alpha_2, \ . \ \forall \mathbf{x} \in R^d \tag{2.2}$$

(2) $V(0) = \min_{\mathbf{x}} V(\mathbf{x})$.

Note that $\pi(\mathbf{x})$ is positive everywhere on R^d, and hence $\{\mathbf{X}^{(n)}\}$ is irreducible and aperiodic and has π as its unique invariant measure.

To represent $\{\mathbf{X}^{(n)}\}$ by using stochastic difference equation, we define $\phi_k (1 \leq k \leq d)$ and $\psi_k (2 \leq k \leq d)$ on R^{d-1} as follows:

For $\mathbf{w} = (w_1, w_2, \cdots, w_{d-1}) \in R^{d-1}$,

$$\begin{aligned} &V(w_1, \cdots, w_{k-1}, \phi_k(\mathbf{w}), w_k, \cdots, w_{d-1}) \\ &= \min_y V(w_1, \cdots, w_{k-1}, y, w_k \cdots, w_{d-1}) \end{aligned}$$

and

$$\begin{aligned} \psi_2(\mathbf{w}) &= \phi_2(\phi_1(\mathbf{w}), w_2, \cdots, w_{d-1}) \\ \psi_3(\mathbf{w}) &= \phi_3(\phi_1(\mathbf{w}), \psi_2(\mathbf{w}), w_3, \cdots, w_{d-1}) \\ &\vdots \\ \psi_d(\mathbf{w}) &= \phi_d(\phi_1(\mathbf{w}), \psi_2(\mathbf{w}), \cdots, \psi_{d-1}(\mathbf{w})). \end{aligned}$$

Let

$$\Psi = (\psi_2, \psi_3, \cdots, \psi_d).$$

Then $\mathbf{X}^{(n+1)} = (X_1^{(n+1)}, \mathbf{Z}^{(n+1)})$ can be denoted by

$$X_1^{(n+1)} = \phi_1(\mathbf{Z}^{(n)}) + \xi^{(n)}$$

$$\mathbf{Z}^{(n+1)} = \Psi(\mathbf{Z}^{(n)}) + \eta^{(n)}.$$

Here $\mathbf{Z}^{(n+1)} = (X_2^{(n+1)}, X_3^{(n+1)}, \cdots, X_d^{(n+1)})$, and $\{\xi^{(n)}\}$ and $\{\eta^{(n)}\}$ are sequences of random vectors in R and R^{d-1}, respectively.

Define $\widehat{V} : R^{d-1} \to R$ by

$$\widehat{V}(\mathbf{z}) = V(\phi_1(\mathbf{z}), \mathbf{z}), \quad \forall \mathbf{z} \in R^{d-1}. \tag{2.3}$$

We can easily show that $\frac{1}{2}\alpha_1|\mathbf{z}|^2 \leq \widehat{V}(\mathbf{z}) \leq \frac{1}{2}\alpha_2|\mathbf{z}|^2$.

We write

$$Pg(\mathbf{x}) := \int g(\mathbf{y})P(\mathbf{x}, d\mathbf{y}), \quad \pi(g) := \int g(\mathbf{y})d\pi(\mathbf{y}).$$

THEOREM 2.1. *Under the assumption (A), the Gibbs sampler $\{\mathbf{X}^{(n)}\}$ is geometrically ergodic.*

PROOF. Define a Lyapunov function $g : R^d \to R$ by

$$g((x_1, \cdots, x_d)) = e^{\lambda \widehat{V}(x_2, \cdots, x_d)} + k, \tag{2.4}$$

where $k \geq 1$ is any constant and $\lambda > 0$ is given later.

Then, for $\mathbf{x}^{(n)} = (x_1^{(n)}, \cdots, x_d^{(n)})$, $\mathbf{z}^{(n)} = (x_2^{(n)}, \cdots, x_d^{(n)})$, we may choose $c_1, c > 0$ and $0 < \gamma < 1$ such that

$$\begin{aligned} Pg(\mathbf{x}^{(n)}) &= E[g(\mathbf{X}^{(n+1)}) \mid \mathbf{X}^{(n)} = \mathbf{x}^{(n)}] \\ &= E[e^{\lambda \widehat{V}(\mathbf{z}^{(n+1)})} + k \mid \mathbf{X}^{(n)} = \mathbf{x}^{(n)}] \\ &\leq c_1 e^{\lambda \widehat{V}(\Psi(\mathbf{z}^{(n)}))} e^{(\frac{\lambda^2}{4\alpha^2} |\nabla V(\phi_1(\Psi(\mathbf{z}^{(n)})), \Psi(\mathbf{z}^{(n)}))|^2)} + k \\ &\leq c_1 e^{\lambda(1+c\lambda)\gamma \widehat{V}(\mathbf{z}^{(n)})} + k. \end{aligned} \tag{2.5}$$

Since the proofs of two inequalities in (2.5) closely follow from those of the theorem 3.5 in Hwang and Sheu (1998), we refer to that paper for details. Now choose λ so that $(1 + c\lambda)\gamma < 1$. If we take $q_0 > 0$ so large that $\lambda(1 - (1 + c\lambda)\gamma)\widehat{V}(\mathbf{z}^{(n)}) > 1 + \ln c_1$ if $|\mathbf{z}^{(n)}| > q_0$, we get

$$Pg(\mathbf{x}^{(n)}) \leq \frac{1}{e} g(\mathbf{x}^{(n)}) + \left(k - \frac{k}{e}\right), \quad |\mathbf{z}^{(n)}| > q_0.$$

Then we may choose r, $\frac{1}{e} < r < 1$ and $q > q_0$ such that

$$(2.6) \qquad Pg(\mathbf{x}) \le rg(\mathbf{x}), \quad |\mathbf{z}| > q$$

and

$$(2.7) \qquad Pg(\mathbf{x}) \le b < \infty, \quad |\mathbf{z}| \le q.$$

Note that $\{\mathbf{X}^{(n)}\}$ is a Feller chain, and therefore every compact set is small. That is, there are constants $\epsilon > 0$, $n_0 \ge 1$, and a probability measure ν such that $P^{(n_0)}(\mathbf{x}, \cdot) \ge \epsilon\nu(\cdot)$, $|\mathbf{x}| \le q$. But, since $P^{(n_0)}(\mathbf{x}, \cdot)$ does not depend on the first coordinate x_1 of $\mathbf{x} = (x_1, \mathbf{z})$, we have $P^{(n_0)}(\mathbf{x}, \cdot) \ge \epsilon\nu(\cdot)$, $|\mathbf{z}| \le q$, which implies that $\{\mathbf{x} = (x_1, \mathbf{z}) | |\mathbf{z}| \le q$ for some $q > 0\}$ is a small set for $\{\mathbf{X}^{(n)}\}$. Hence from (2.6) and (2.7), the conclusion follows. □

(2.6) together with (2.7) can be written in the form of

$$(2.8) \qquad Pg(\mathbf{x}) \le rg(\mathbf{x}) + bI_C,$$

where $b < \infty$ and I_C is the indicator function of C.

Moreover, (2.8) is equivalent to the following statement (Meyn and Tweedie (1993), Chapter 15); There are $M < \infty$ and $\rho < 1$ such that

$$(2.9) \qquad \|P^{(n)}(\mathbf{x}, \ ) - \pi\|_g \le Mg(\mathbf{x})\rho^n,$$

where for any signed measure μ, the g−norm is defined as

$$\|\mu\|_g = \sup_{|f| \le g} \left| \int f(\mathbf{y})\mu(d\mathbf{y}) \right|.$$

One of the important results which follows from the proof of geometric ergodicity is a (functional) central limit theorem. We consider the functional central limit theorem for

$$(2.10) \qquad Y_n(t) = \frac{1}{\sqrt{n}} \sum_{k=0}^{[nt]} (f(\mathbf{X}^{(k)}) - \pi(f)), \quad t \ge 0.$$

which is studied by many authors (see, e.g. Roberts and Tweedie (1996), Tierney (1994), Hobert and Geyer (1998) etc.).

Let $\|\cdot\|_2$ denote the L^2 norm on R^d.

Next two lemmas are applied to find a class of functions for which functional central limit theorems holds.

LEMMA 2.1 (Gordin and Lifšic (1978)). *For Harris ergodic chain, suppose that f is in the range of $I-P$, say $f=(I-P)h$ for some $h \in L^2(\pi)$. Then $Y_n(t)$ in (2.10) converges to a Brownian motion with mean 0 and variance parameter given by $\sigma(f)^2 = \|h\|_2^2 - \|Ph\|_2^2$ for any initial distribution.*

LEMMA 2.2. *If $\sum_{n=0}^{\infty} \|P^n(f-\pi(f))\|_2 < \infty$, then $f-\pi(f)$ belongs to the range of $P-I$.*

PROOF. Let $h = -\sum_{n=0}^{\infty} P^n(f-\pi(f))$ and apply $P-I$ to the right side of the equation. □

THEOREM 2.2. *If $g \in L^2(\pi)$ and $|f| \le g$, then the functional central limit theorem holds for f.*

PROOF. Suppose that $|f| \le g$. Then, from (2.9), $|P^n f(\mathbf{x}) - \pi(f)| \le Mg(\mathbf{x})\rho^n$ and then

$$\|P^n f(\mathbf{x}) - \pi(f)\|_2^2 = \int |P^n f(\mathbf{x}) - \pi(f)|^2 \pi(d\mathbf{x}) \le M^2 \rho^{2n} \int (g(\mathbf{x}))^2 \pi(d\mathbf{x}).$$

Hence if $g \in L^2(\pi)$, then for any f with $|f| \le g$, $\sum_{n=0}^{\infty} \|P^n(f-\pi(f))\|_2 < \infty$. By above lemmas, functional central limit theorem holds for f. □

COROLLARY 2.1. *Functional central limit theorem holds for any f such that $f^2 \le g$.*

PROOF. From (2.8), we have

$$Pg^{\frac{1}{2}} \le \lambda^{\frac{1}{2}} g^{\frac{1}{2}} + b^{\frac{1}{2}} I_C. \tag{2.11}$$

Note that $\pi(g) < \infty$. If $|f| \le g^{\frac{1}{2}}$, i.e., $f^2 \le g$, by (2.11) and theorem (2.2), functional central limit theorem holds for f. □

COROLLARY 2.2. *Functional central limit theorem holds for every bounded measurable function f.*

PROOF. Let $|f| \le C$ for some $C < \infty$. If we take $k = C^2$ in g in (2.4), then $f^2 \le g$. Apply corollary 2.1 to obtain the result. □

References

[1] Athreya, K. B., Doss, H. and Sethuraman, J., *On the convergence of the Markov chain simulation method,* The Annals of Statistics **24** (1996), no. 1, 69-100.

[2] Chan, K. S., *Asymptotic Behavior of the Gibbs sampler*, J. American Statistical Association **88** (1993), 320-326.

[3] Gelfand, A. E. and Smith, A. F. M., *Sampling-based approaches to calculating marginal densities*, J. American Statistical Association **85** (1990), 398-409.

[4] Geyer, C. J., *Practical Markov chain Monte Carlo*, Statistical Science **7** (1992), no. 4, 473-511.

[5] Glynn, P. W. and Meyn, S. P., *A Lyapunov bound for solutions of the poisson equation*, The Annals of Probability **24** (1996), no. 2, 916-931.

[6] Gordin, M. I. and Lifšic, B. A., *The central limit theorem for stationary Markov processes*, Soviet Math. Dokl. **19** (1978), 392-394.

[7] Hobert, J. P. and Casella, G., *Functional compatibility, Markov chains, and Gibbs sampling with improper Posteriors*, J. Computational and Graphical Statistics **7** (1998), no. 1, 42-60.

[8] Hobert, J. P. and Geyer, C. J., *Geometric ergodicity of Gibbs and block Gibbs sampler for hierarchical random effects model*, J. Multivariate Analysis **67** (1998), 414-430.

[9] Hwang, C. and Sheu, S., *On the geometrical convergence of Gibbs sampler in R^d*, J. Multivariate Analysis **66** (1998), 22-37.

[10] Mengersen, K. L. and Tweedie, R. L., *Rates of convergence of the Hastings and Metropolis algorithms*, The Annals of Statistics **24** (1996), no. 1, 101-121.

[11] Meyn, S. P. and Tweedie, R. L., *Markov chains and stochastic stability*, London, springer-Verlag, 1993.

[12] Roberts, G. O. and Rosental, J. S., *Markov chain Monte Carlo: Some practical implications of theoretical results*, Canadian Journal of Statistics **26** (1998), no. 1, 5-31.

[13] Roberts, G. O. and Tweedie, R. L., *Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms*, Biometrica **83** (1996), no. 1, 95-110.

[14] Tierney, L., *Markov chains for exploring posterior distributions*, The Annals of Statistics **22** (1994), 1701-1762.

Department of Statistics
Ewha Womans University
Seoul 120-750, Korea
E-mail: oslee@mm.ewha.ac.kr