UNIVERSITY OF WISCONSIN

DEPARTMENT OF BIOSTATISTICS AND MEDICAL INFORMATICS

Technical Report #128

May 1998

MONTE CARLO ERROR ESTIMATION FOR MULTIVARIATE MARKOV CHAINS

Michael R. Kosorok, PhD

MADISON, WISCONSIN

Monte Carlo error estimation for multivariate Markov chains

Michael R. Kosorok [1]

ABSTRACT

In this paper, the conservative Monte Carlo error estimation methods and theory developed in Geyer (1992a) are extended from univariate to multivariate Markov chain applications. A small simulation study demonstrates the feasibility of the proposed estimators.

Keywords: Autocovariance estimators, Greatest convex minorants, Markov chain Monte Carlo, Metropolis-Hastings algorithm, Stationary processes, Window estimators.

[1]Michael R. Kosorok is Associate Professor, Departments of Statistics and Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI 53706. Michael R. Kosorok's research was paid in part by a University of Wisconsin Graduate School grant and by National Cancer Institute grant CA75142.

1 Introduction

The method of Markov chain Monte Carlo (MCMC) described by Metropolis et al (1953) and Hastings (1970), is becoming an increasingly popular method for generating stationary stochastic processes with probability densities known only up to a constant of proportionality (see Besag and Green, 1993; Besag et al, 1995; Chib and Greenberg, 1995; Geyer, 1992a; Gilks et al, 1993; and Smith and Roberts, 1993). Assume that one of these methods has been chosen to generate the irreducible multivariate Markov chain $\{\beta_i, i \geq 1\}$, with stationary distribution of interest P, with $\beta_1 \in \Re^p$ for finite p. Interest often centers in estimating $\mu \equiv \int_{\Re^p} g(x)dP(x)$ for some P-measurable g (possibly vector valued). μ can be estimated by

$$\hat{\mu}_n \equiv n^{-1} \sum_{i=1}^{n} g(\beta_i),$$

from a subchain of length n. Generally, an MCMC approach can be chosen which ensures that the chain $\{\beta_i, i \geq 1\}$ is stationary, irreducible, and reversible: we will assume this is the case throughout. Under these assumptions, the chain $\{g(\beta_i), i \geq 1\}$ is also stationary, irreducible, and reversible. Thus, without loss of generality, we will assume throughout that g is the identity mapping; hence, $\mu \equiv \int_{\Re^p} x dP(x)$ and

$$\hat{\mu}_n \equiv n^{-1} \sum_{i=1}^{n} \beta^i.$$

The goal of the present paper is to obtain a conservative estimate of the Monte Carlo error $n \times \text{var}[\hat{\mu}_n]$. This will be accomplished by generalizing the conservative univariate approach to Markov chain Monte Carlo error estimation given in Geyer (1992a) to the multivariate setting. As Geyer points out, the motivation for using a conservative (ie., possibly upward biased) variance estimator is that, in many applications, overestimates are less of a problem than underestimates. This is certainly true when the main purpose of Monte Carlo error estimation is to help determine when a Markov chain is sufficiently long for conducting inference on the stationary distribution, as discussed, for example, by Tierney (1992). A comparison of the performance of the univariate versions of these estimators with other kinds of estimators, such as window or batched means estimators, is presented in Geyer (1992a). As pointed out by Raftery and Lewis (1992), the Monte Carlo error estimators proposed by Geyer (1992a) can sometimes perform poorly compared to other window estimators—such as the Tukey-Hanning window estimator (Priestley, 1981, page 443)—when the chain is not stationary. However, a key advantage Geyer's estimators have is that they are more automatic and can perform quite well for chains which actually are in equilibrium (Geyer, 1992b).

In the next section, we present several results which help formulate the general strategy behind our proposed estimators (proofs of results are generally postponed until the last section). The three estimators are then presented in Section 3 with a brief simulation study given in Section 4. The paper concludes with technical proofs given in Section 5.

2 Multivariate Markov chains and matrix maximums

Before proceeding, we need several definitions. Let

$$G_j \equiv \frac{E\left[\{\beta_1 - \mu\}\{\beta_{1+j} - \mu\}^T\right] + E\left[\{\beta_{1+j} - \mu\}\{\beta_1 - \mu\}^T\right]}{2}$$

be the symmetrized j'th autocovariance, where $\mu = E[\beta_1]$ and $j \geq 0$; and define $\Gamma_j \equiv G_{2j} + G_{2j+1}$, also for $j \geq 0$. In addition, for a subchain of length n, define the estimators

$$\widehat{G}_j \equiv \frac{\sum_{i=1}^{n-j}\left[\{\beta_i - \hat{\mu}_n\}\{\beta_{i+j} - \hat{\mu}_n\}^T + \left\{\beta^{i+j} - \hat{\mu}_n\right\}\{\beta_i - \hat{\mu}_n\}^T\right]}{2n},$$

and $\widehat{\Gamma}_j = \widehat{G}_{2j} + \widehat{G}_{2j+1}$. For a real symmetric matrix A, let $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the smallest and largest eigenvalues, respectively, of A.

The following generalization of Theorem 3.1 of Geyer (1992a) gives us precise information on the structure of these autocovariance terms in our setting:

THEOREM 1 *For a stationary, irreducible, reversible Markov chain in $\Re^p$ with symmetrized autocovariance terms G_j and Γ_j, as defined above for $j \geq 0$, we have that $\lambda_{min}(\Gamma_j)$, $\lambda_{max}(\Gamma_j)$, $\lambda_{min}(\Gamma_j - \Gamma_{j+1})$, and $\lambda_{max}(\Gamma_j - \Gamma_{j+1})$ are strictly positive and strictly decreasing; that $\lambda_{max}(\Gamma_j)$ is strictly convex in j; and that $\lambda_{min}(\Gamma_j - 2\Gamma_{j+1} + \Gamma_{j+2}) > 0$, for $0 \leq j < \infty$.*

The above result essentially dictactes the three estimators we propose. The first estimator essentially adds up autocovariances terms from $\widehat{G}_0$ through $\widehat{G}_{2\hat{m}+1}$, where $\hat{m}$ is the largest value of j such that $\lambda_{min}\left(\widehat{\Gamma}_j\right) \geq 0$, corresponding to Geyer's *initial positive sequence estimator*. To obtain multivariate estimators corresponding to Geyer's *initial monotone sequence estimator* and *initial convex sequence estimator*, we will need a suitable method of finding the greatest decreasing minorant of a collection of real symmetric matrices as well as a method of obtaining the greatest convex minorant of a decreasing sequence of real symmetric matrices. Although details of accomplishing this will be presented in the next section, the method we have developed for maximizing a given collection of real symmetric matrices is so crucial to this process that we will present it now.

Let $\mathbf{M}^p$ be the space of real symmetric $p \times p$ matrices. For $A \in \mathbf{M}^p$, let A^+ be the smallest positive semidefinite matrix B such that $B \geq A$. It is trivial to show that if $M\Lambda M^T$ is the canonical decomposition of A (where superscript T denotes transpose) then $A^+ = M\Lambda^* M^T$, where Λ^* is the diagonal matrix Λ with negative eigenvalues replaced by zero, is uniquely defined. A^+ is essentially the positive part of A. We similarly define $A^- \equiv [-A]^+$ to be the negative part of A. Clearly, $A = A^+ - A^-$ and the spaces spanned by the columns of A^+ and A^- are orthogonal.

For $A, B \in \mathbf{M}^p$, let $A \vee B \equiv A + [B - A]^+$ be the matrix maximum of A and B. That this is well defined is contained in the following lemma:

LEMMA 1 *For $A, B \in \mathbf{M}^p$, $A \vee B$ is the unique smallest matrix $\mathbf{C} \in \mathbf{M}^p$ such that $\mathbf{C} \geq A$ and $\mathbf{C} \geq B$.*

For a collection of matrices $A_1, \ldots, A_k \in \mathbf{M}^p$, define the matrix maximum recursively as follows:

$$\max\{A_1, \ldots, A_k\} \equiv \max\{A_1, \ldots, A_{k-1}\} \vee A_k$$

and $\max\{A_1, A_2\} \equiv A_1 \vee A_2$:

LEMMA 2 *For $A_1, \ldots, A_k \in \mathbf{M}^p$, $\max\{A_1, \ldots, A_k\}$ is the unique smallest matrix $\mathbf{C} \in \mathbf{M}^p$ such that $\mathbf{C} \geq A_i$ for $i = 1 \ldots k$.*

An obvious consequence of Lemma 2 is that the matrix maximum is independent of the order of the matrices given in the maximization expression.

We can analogously define the matrix minimums

$$A \wedge B \equiv -\{[-A] \vee [-B]\}$$

and

$$\min\{A_1, \ldots, A_k\} \equiv -\max\{-A_1, \ldots, -A_k\},$$

for $A, B, A_1, \ldots, A_k \in \mathbf{M}^p$. Clearly, these will have uniqueness properties corresponding to the results given in Lemmas 1 and 2 for matrix maximums.

3 Three Monte Carlo error estimators

We now formulate three potentially useful estimators of the Monte Carlo error. Define

$$\hat{m} \equiv \inf\left\{j \geq 1 : \lambda_{min}\left(\widehat{\Gamma}_j\right) \leq 0\right\} - 1.$$

The first Monte Carlo error variance estimator, the *initial positive sequence estimator*, is

$$\widehat{H}_1 \equiv -\widehat{G}_0 + 2\sum_{j=0}^{\hat{m}} \widehat{\Gamma}_j.$$

The second estimator, the *initial monotone sequence estimator*, is computed by forming the greatest monotone matrix minorant of the sequence $\left\{\widehat{\Gamma}_0, \ldots, \widehat{\Gamma}_{\hat{m}}\right\}$. This is accomplished by computing $\widehat{\Gamma}_0^{(2)} = \widehat{\Gamma}_0$ and $\widehat{\Gamma}_j^{(2)} = \widehat{\Gamma}_{j-1}^{(2)} \wedge \widehat{\Gamma}_j$ for $j = 1 \ldots \hat{m}$ (provided $\hat{m} > 0$) and then taking

$$\widehat{H}_2 \equiv -\widehat{G}_0 + 2\sum_{j=0}^{\hat{m}} \widehat{\Gamma}_j^{(2)}$$

to be the second estimator. By Lemma 2, the sequence $\left\{\widehat{\Gamma}_0^{(2)}, \ldots, \widehat{\Gamma}_{\hat{m}}^{(2)}\right\}$ is the unique largest monotome decreasing sequence of matrices less than or equal to $\left\{\widehat{\Gamma}_0, \ldots, \widehat{\Gamma}_{\hat{m}}\right\}$.

Before defining the third estimator, the *initial convex sequence estimator*, we first need to define the greatest convex matrix minorant of a descending sequence of matrices. For a descending collection of $k \geq 3$ matrices, $\{A_1, \ldots, A_k\}$, let

$$\text{cvx}\{A_1, \ldots, A_k\} \equiv \{A_1^*, \ldots, A_k^*\}$$

be the greatest convex matrix minorant, determined recursively as follows: $A_1^* = A_1$; for i such that $1 < i < k$,

$$A_i^* = A_{i-1}^* - \max_{j:i \leq j < k}\left[\frac{A_{i-1}^* - A_j}{i - j + 1}\right],$$

and $A_k^* = A_k$. We have the following lemma:

Lemma 3 *Let* $\{A_1, \ldots, A_k\}$ *be a descending sequence of* $k \geq 3$ *matrices in* $\mathbf{M}^p$. *Then*

$$cvx\{A_1, \ldots, A_k\} \equiv \{A_1^*, \ldots, A_k^*\}$$

is the unique greatest convex matrix minorant of $\{A_1, \ldots, A_k\}$, *in the sense that if* $\{B_i \in \mathbf{M}^p, 1 \leq i \leq k\}$ *is any other* $\mathbf{M}^p$ *matrix collection with* $\lambda_{min}\{B_{i-1} - 2B_i + B_{i+1}\} \geq 0$ *for* $1 < i < k$ *and with* $B_i \leq A_i$ *for* $1 \leq i \leq k$, *then* $B_i \leq A_i^*$ *for* $1 \leq i \leq k$.

The third estimator is now obtained by computing

$$\left\{\widehat{\Gamma}_0^{(3)}, \ldots, \widehat{\Gamma}_{\hat{m}}^{(3)}\right\} \equiv \text{cvx}\left\{\widehat{\Gamma}_0^{(2)}, \ldots, \widehat{\Gamma}_{\hat{m}}^{(2)}\right\}$$

and setting

$$\widehat{H}_3 \equiv -\widehat{G}_0 + 2\sum_{i=0}^{\hat{m}} \widehat{\Gamma}_i^{(3)}.$$

That this is a unique and well-defined *initial convex sequence estimator* follows from Lemma 3.

The following theorem, which is a multivariate generalization of Theorem 3.2 of Geyer (1992a), shows that the above three variance estimators are asymptotically conservative for the true Monte Carlo covariance $H_0 \in \mathbf{M}^p$:

THEOREM 2 *Under the assumptions of Theorem 1, and provided* $\|H_0\| < \infty$ *(where for a matrix* C*,* $\|C\|$ *is the maximum in absolute value over all its elements), we have for any real vector* $u \in \Re^p$*, that*

$$\liminf_{n\to\infty} u^T \widehat{H}_1 u \geq \liminf_{n\to\infty} u^T \widehat{H}_2 u \geq \liminf_{n\to\infty} u^T \widehat{H}_3 u \geq u^T H_0 u.$$

4 A small simulation study

A small simulation study was conducted to explore briefly the previously proposed estimators and compare them to the "truncated periodogram" window estimator (Priestly, 1981, p. 437) with window width $n^{2/3}$:

$$\widehat{H}_0 \equiv \widehat{G}_0 + 2\sum_{i=1}^{n^{2/3}} \widehat{G}_i.$$

The power 2/3 was chosen to ensure that the $\widehat{H}_0$ would utilize at least as many autocovariance terms as the other estimators considered. A simple $\Re^3$ AR(1) process was utilized with lag-one autocorrelation $\rho = 0.98$ and with independent standard normal (trivariate) marginal. The true Monte Carlo error for this process is

$$H_0 \equiv \begin{bmatrix} 99 & 0 & 0 \\ 0 & 99 & 0 \\ 0 & 0 & 99 \end{bmatrix}$$

($99 = 1 + 2\frac{\rho}{1-\rho} = 1 + 2\frac{0.98}{0.02}$). 20 independent replicates of this process with length 10,000 were generated; and, for each replicate, the estimators $\widehat{H}_0$, $\widehat{H}_1$, $\widehat{H}_2$, and $\widehat{H}_3$ were computed. The absolute error for each estimate was computed by determining the largest absolute value of the eigenvalues of the matrix resulting from subtracting H_0 from the estimate.

The resulting means of each of the estimators, $\widehat{H}_0$, $\widehat{H}_1$, $\widehat{H}_2$, and $\widehat{H}_0$, are as follows:

$$\overline{H}_0 = \begin{bmatrix} 104.71 & 5.75 & -10.97 \\ 5.75 & 100.32 & -0.59 \\ -10.97 & -0.59 & 77.40 \end{bmatrix}; \qquad \overline{H}_1 = \begin{bmatrix} 92.43 & 1.68 & -2.99 \\ 1.68 & 98.41 & -3.29 \\ -2.99 & -3.29 & 87.80 \end{bmatrix};$$

$$\overline{H}_2 = \begin{bmatrix} 91.74 & 1.74 & -2.74 \\ 1.74 & 98.20 & -3.34 \\ -2.74 & -3.34 & 87.39 \end{bmatrix}; \qquad \overline{H}_3 = \begin{bmatrix} 91.17 & 1.79 & -2.94 \\ 1.79 & 97.48 & -3.15 \\ -2.94 & -3.15 & 86.55 \end{bmatrix}.$$

The resulting mean absolute error (with standard deviation in parentheses) for each of the estimators is: 64.71 (26.10) for $\widehat{H}_0$; 36.68 (9.24) for $\widehat{H}_1$; 36.79 (9.39) for $\widehat{H}_2$; and 36.72 (9.30) for $\widehat{H}_3$. Using the Wilcoxon signed-rank test, the absolute error for $\widehat{H}_0$ is significantly larger than any $\widehat{H}_j$, $j = 1 \ldots 3$, with a two-sided p-value less than 0.0001. On the other hand, none of the differences among the $\widehat{H}_j$, $j = 1 \ldots 3$, differed significantly from zero using the same procedure: all two-sided p-values were larger than 0.25. However, in all cases, the ordering

$$\widehat{H}_1 \geq \widehat{H}_2 \geq \widehat{H}_3$$

was preserved. For each of the 20 simulations, the largest eigenvalues of the differences $\widehat{H}_1 - \widehat{H}_2$, $\widehat{H}_1 - \widehat{H}_3$, and $\widehat{H}_2 - \widehat{H}_3$ were computed: the means (with standard deviations in parentheses) of these largest eigenvalues are 1.04 (2.45) for $\widehat{H}_1 - \widehat{H}_2$, 1.96 (2.84) for $\widehat{H}_1 - \widehat{H}_3$, and 1.50 (2.10) for $\widehat{H}_2 - \widehat{H}_3$.

Close examination revealed that the improvement in accuracy of the $\widehat{H}_j$ ($j = 1 \ldots 3$) estimators was due to a substantial reduction in both bias and variability. It is possible that more substantial differences in performance between the $\widehat{H}_j$ estimators for $j = 1 \ldots 3$ may emerge for Markov chains more complex than the simple AR(1) process studied here. An important caveat to mention is that, although the asymptotics and this small simulation study are encouraging, there are no constraints ensuring positive definiteness of the proposed estimators for finite chains; thus care may be needed in certain applications (of course, this is an issue for window estimators in general). However, the main purpose of this simulation study has been to establish feasibility of the proposed estimators; and other studies will be needed for more definitive comparisons. These estimators are also quite easy and fast to compute: a macro for this purpose written in the S-Plus (StatSci: A Division of MathSoft, Inc., Seattle, WA) statistical computing language is available from the author.

5 Proofs

Before giving the proof of Theorem 1, we need the following lemma:

LEMMA 4 *If $A, B \in \mathbf{M}^p$, and if $v^T A v > v^T B v$ for all unit vectors $v \in \Re^p$, then $\lambda_{min}(A) > \lambda_{min}(B)$ and $\lambda_{max}(A) > \lambda_{max}(B)$.*

Proof. Let u be an eigenvector of B corresponding to $\lambda_{max}(B)$, then

$$\lambda_{max}(A) \geq u^T A u > u^T B u = \lambda_{max}(B).$$

If we next let u be an eigenvector of A corresponding to $\lambda_{min}(A)$, then

$$\lambda_{min}(A) = u^T A u > u^T B u \geq \lambda_{min}(B).\square$$

Proof of Theorem 1. Assume the column dimension of Γ_0 is p. By Theorem 3.1 of Geyer (1992a), $v^T \Gamma_j v$, $v^T (\Gamma_j - \Gamma_{j+1}) v$, and $v^T (\Gamma_j - 2\Gamma_{j+1} + \Gamma_{j+2}) v$ are strictly positive for every unit vector $v \in \Re^p$ and all $j \geq 0$; and we thus get strict positivity for all of the eigenvalues mentioned in Theorem 1. Geyer's theorem also implies that both $v^T \Gamma_j v > v^T \Gamma_{j+1} v$ and $v^T (\Gamma_j - \Gamma_{j+1}) v > v^T (\Gamma_{j+1} - \Gamma_{j+2}) v$ for every unitve vector $v \in \Re^p$ and all $j \geq 0$; and Lemma 4 now implies the strictly decreasing part of Theorem 1. The only part remaining is to prove that $\lambda_{max}(\Gamma_j)$ is strictly convex in $j \geq 0$. By preceding arguments,

$$v^T \Gamma_j v - v^T \Gamma_{j+1} v > v^T \Gamma_{j+1} v - v^T \Gamma_{j+2} v,$$

for every unit vector $v \in \Re^p$, which implies

$$\lambda_{max}(\Gamma) - v^T \Gamma_{j+1} v > v^T \Gamma_{j+1} v - \lambda_{max}(\Gamma_{j+2}),$$

also for every unit vector $v \in \Re^p$. In particular, choose v to be an eigenvector corresponding to $\lambda_{max}(\Gamma_{j+1})$, and the result follows.$\square$

Proof of Lemma 1. Let C be such that $C \geq A$ and $C \geq B$, then both $C - A$ and $C - B$ must be positive semidefinite and $C = A + D_1 = B + D_2$ where both D_1 and D_2 are positive semidefinite. However,

$$\begin{aligned} A + D_1 &= B - [B - A] + D_1 \\ &= B - [B - A]^+ + [B - A]^- + D_1, \end{aligned}$$

which implies that $D_1 - [B - A]^+ + [B - A]^-$ is positive semidefinite. Now suppose there exists $u \in \Re^p$ such that

$$u^T \left\{ D_1 - [B - A]^+ \right\} u < 0;$$

but this implies

$$u^T D_1 u - u_*^T [B-A]^+ u_* < 0,$$

where u_* is the projection of u onto the space spanned by the columns of $[B-A]^+$. However, since $u^T D_1 u \geq u_*^T D_1 u_*$ by the positive semidefiniteness of D_1, we have that

$$u_*^t \left\{ D_1 - [B-A]^+ \right\} u_* < 0$$

and thus

$$u_*^t \left\{ D_1 - [B-A]^+ + [B-A]^- \right\} u_* < 0$$

by the orthogonality of the space spanned by the columns of $[B-A]^+$ and the space spanned by the columns of $[B-A]^-$. But this is a contradiction of the positive definiteness of $D_1 - [B-A]^+ + [B-A]^-$ and thus $D_1 \geq [B-A]^+$ and $A \vee B$ is the smallest $C \in \mathbf{M}^p$ satisfying $C \geq A$ and $C \geq B$. The uniqueness follows after using the above arguments to establish that $B \vee A \leq A \vee B$.□

Proof of Lemma 2. This is a simple consequence of Lemma 1.

Proof of Lemma 3. First $A_i^* \leq A_i$ for $1 \leq i \leq k$. Furthermore, for $1 < i \leq j \leq k$,

$$\frac{A_{i-1}^* - A_j}{j-i+1} \leq A_{i-1}^* - A_j \leq A_{i-1}^* - A_k,$$

thus $A_i^* \geq A_k$ for $1 \leq i \leq k$ and $A_k^* = A_k$. In addition, for i such that $1 < i < k$,

$$\min_{j:1\leq j<i} \left[\frac{A_j^* - A_i^*}{j-i} \right] \geq \max_{j:i<j\leq k} \left[\frac{A_i^* - A_j^*}{i-j} \right]$$

by construction; thus $\lambda_{min} \left\{ A_{i-1}^* - 2A_i^* + A_{i+1}^* \right\} \geq 0$ for all i such that $1 < i < k$, and we have established that $\{A_1^*, \ldots, A_k^*\}$ is a convex minorant of $\{A_1, \ldots, A_k\}$.

Now, let $\{B_1, \ldots, B_k\}$ be any other convex minorant of $\{A_1, \ldots, A_k\}$. We must show that for i such that $1 < i \leq k$,

$$B_i \leq B_{i-1} - \max_{j:i\leq j\leq k} \left[\frac{B_{i-1} - A_j}{j-i+1} \right].$$

If this were not true, there would exist $u \in \Re^p$ such that

$$u^T B_i u > u^T \left\{ \min_{j:i\leq j\leq k} \left[\frac{(j-i)B_{i-1} + A_j}{j-i+1} \right] \right\} u$$

which implies

$$u^T B_i u > u^t \left\{ \frac{(j-i)B_{i-1} + A_j}{j-i+1} \right\} u$$

for some j such that $i < j \leq k$; but this implies

$$u^T B_i u > u^t \left\{ \frac{(j-i)B_{i-1} + B_j}{j-i+1} \right\} u,$$

since $B_j \leq A_j$, and convexity will fail. Hence

$$B_i \leq B_{i-1} - \max_{j:i\leq j\leq k} \left[\frac{B_{i-1} - A_j}{j-i+1} \right]$$

for all i such that $1 < i \leq k$. Clearly, $B_1 \leq A_1^*$. For any i such that $1 < i \leq k$, if $B_{i-1} \leq A_{i-1}^*$, then

$$\begin{aligned} B_i &\leq B_{i-1} - \max_{j:i\leq j\leq k} \left[\frac{B_{i-1} - A_j}{j-i+1} \right] \\ &= \min_{j:i\leq j\leq k} \left[\frac{(j-i)B_{i-1} - A_j}{j-i+1} \right] \\ &\leq \min_{j:i\leq j\leq k} \left[\frac{(j-i)A_{i-1}^* - A_j}{j-i+1} \right] \\ &= A_i^*, \end{aligned}$$

and the proof is complete.□

Before giving the proof of Theorem 2, we will need the following additional lemma:

LEMMA 5 *For* $A, E \in \mathbf{M}^p$, $\left\| [A+E]^+ - A^+ \right\| \leq p^2 \|E\|$.

Proof. Note that

$$\begin{aligned} [A+E]^+ - A^+ &\leq A^+ + E^+ - A^+ \\ &= E^+ \end{aligned}$$

and thus for any $v \in \Re^p$, $v^T \left\{ [A+E]^+ - A^+ \right\} v \leq p \|E\|$. Now set $B = A + E$ and note that

$$\begin{aligned} [A+E]^+ - A^+ &= B^+ - [B-E]^+ \\ &\geq B^+ - B^+ - [-E]^+ \\ &= -E^- \end{aligned}$$

which implies for any $v \in \Re^p$ that $v^T \left\{ [A+E]^+ - A^+ \right\} v \geq -p \|E\|$, and the result follows.□

Proof of Theorem 2. Let the number of columns of H_0 be p. For every $\epsilon > 0$, there exists an $m_0 < \infty$ such that $\left\| \sum_{i=m_0+1}^{\infty} \Gamma_i \right\| < \epsilon$; also, for n large enough, $\widehat{\mathbf{\Gamma}}_i$ satisfies the eigenvalue conditions for Γ_i given in Theorem 1 for $0 \leq i \leq m_0$. Thus with high probability $\widehat{H}_j = E_j - \widehat{G}_0 + 2\sum_{i=0}^{m_0} \widehat{\Gamma}_i^{(j)}$, where E_j is positive semidefinite, for $j = 1, 2, 3$, and where $\widehat{\Gamma}_i^{(1)} \equiv \widehat{\Gamma}_i$ and both $\widehat{\Gamma}_i^{(2)}$ and $\widehat{\Gamma}_i^{(3)}$ are as defined in Section 3, for all $i \geq 0$. It is clear by Lemma 5 that

$$\lim_{n\to\infty} \left\{ \left\| \widehat{G}_0 - G_0 \right\| + \max_{0 \leq i \leq m_0} \left\| \widehat{\Gamma}_i^{(j)} - \Gamma_i \right\| \right\} = 0$$

in probability, for $j = 1 \ldots 3$. We then have for any unit vector $v \in \Re^p$ that

$$\begin{aligned}
\liminf_{n\to\infty} v^T \left(\widehat{H}_j - H_0 \right) v &= \liminf_{n\to\infty} v^T \left(-\widehat{G}_0 + G_0 + 2 \sum_{i=0}^{m_0} \left[\widehat{\Gamma}_i^{(j)} - \Gamma_i \right] + E_j - 2 \sum_{i=m_0+1}^{\infty} \Gamma_i \right) v \\
&= \liminf_{n\to\infty} v^T E_j v - 2 v^T \left(\sum_{i=m_0+1}^{\infty} \Gamma_i \right) v \\
&\geq -2p\epsilon.
\end{aligned}$$

The result now follows since ϵ is arbitrary.□

ACKNOWLEDGMENTS

The author would like to thank Jeffrey A. Douglas and Michael A. Newton for many stimulating discussions on this topic.

REFERENCES

Besag, J., and Green, P. J. (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society*, Ser. B, **55**, 25–37.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3–66.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327–335.

Geyer, C. J. (1992a). Practical Markov chain Monte Carlo. *Statistical Science*, **7**, 473–483.

Geyer, C. J. (1992b). Rejoinder on "Practical Markov chain Monte Carlo." *Statistical Science*, **7**, 502–503.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J. (1993). Modeling complexity: Applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society*, Ser. B, **55**, 39–52.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, New York.

Raftery, A. E., and Lewis, S. M. (1992). Comment on "Practical Markov chain Monte Carlo." *Statistical Science*, **7**, 493–497.

Smith, A. F. M., and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, Ser. B., **55**, 3–23.

Tierney, L. (1992). Comment on "The Gibbs sampler and Markov chain Monte Carlo." *Statistical Science*, **7**, 499–501.