Multivariate Initial Sequence Estimators in Markov Chain Monte Carlo

Ning Dai
School of Statistics
University of Minnesota
daixx224@umn.edu

Galin L. Jones *
School of Statistics
University of Minnesota
galin@umn.edu

March 22, 2017

Abstract

Markov chain Monte Carlo (MCMC) is a simulation method commonly used for estimating expectations with respect to a given distribution. We consider estimating the covariance matrix of the asymptotic multivariate normal distribution of a vector of sample means. Geyer [9] developed a Monte Carlo error estimation method for estimating a univariate mean. We propose a novel multivariate version of Geyer's method that provides an asymptotically valid estimator for the covariance matrix and results in stable Monte Carlo estimates. The finite sample properties of the proposed method are investigated via simulation experiments.

*Research supported by the National Institutes of Health and the National Science Foundation.

1 Introduction

Many distributions encountered in modern applications are intractable in the sense that it is difficult to calculate expectations without resorting to simulation-based methods. If it is difficult to simulate independent realizations from the target distribution, then it is natural to turn to Markov chain Monte Carlo (MCMC). An MCMC experiment consists of generating a realization of an irreducible Markov chain having the distribution of interest as its stationary distribution [22, 25]. The simulated data may then be used to estimate a vector of means associated with the stationary distribution. The reliability of this estimation can be assessed by forming asymptotically valid confidence regions for the means of the stationary distribution [6, 7, 9, 18, 19, 28]. (There is a simliar approach to quantile estimation [3].) The confidence regions are based on estimating the covariance matrix in a multivariate Markov chain central limit theorem (CLT). We propose and study a novel method for estimating this covariance matrix.

Estimating the covariance matrix has been mostly ignored in the MCMC literature until recently. Vats et al. [28] and Vats et al. [29] studied non-overlapping batch means and spectral methods, respectively, and found that these estimators often underestimate the size of the confidence regions and overestimate the effective sample size unless the Monte Carlo sample sizes are enormous. Kosorok [21] proposed an estimator that is closer in spirit to ours than the spectral and batch means methods, but we will see later that it typically overestimates the effective sample size, resulting in overconfidence in the reliability of the simulation. We propose alternative estimators of the covariance matrix that require weaker mixing conditions on the Markov chain and weaker moment conditions on the function of interest than those required by batch means and spectral methods. Specifically, our method applies as long as a Markov chain CLT holds and detailed balance is satisfied, which is not enough to guarantee the asymptotic validity of batch means or spectral methods. We show that the proposed estimators are asymptotically valid and study their empirical performance. The problem we consider will now be described more formally.

Let F be a distribution having support $\mathcal{X}$ and if $p \geq 1$, let $g : \mathcal{X} \to \mathbb{R}^p$ be F-integrable and set

$$\mu := \mathrm{E}_F \{g(X)\} = \int_{\mathcal{X}} g(x)\, F(dx)\,.$$

Also let $\Phi = \{X_0, X_1, X_2, \ldots\}$ be a Harris ergodic–namely, irreducible, aperiodic and Harris recurrent–Markov chain having invariant distribution F. By averaging the function over a realization of Φ, estimation of μ is straightforward since, with probability 1,

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \to \mu \quad \text{as} \quad n \to \infty\,.$$

The Markov chain strong law justifies the use of MCMC but provides no information about the quality of estimation or how large the simulation size n should be. More specifically, additional information is needed to answer either of the following two questions.

1. Given a pre-specified run length n, how reliable is μ_n as an estimate of μ? Specifically, how do we construct a confidence region for μ?

2. How large should the simulation size n be to ensure a reliable estimate of μ?

We can address these issues through the approximate sampling distribution of the *Monte Carlo error*, $\mu_n - \mu$. A Markov chain CLT exists when there is a positive definite matrix Σ such that, as $n \to \infty$,

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{d} \mathcal{N}_p(0, \Sigma)\,. \tag{1}$$

See Jones [17] and Roberts and Rosenthal [26] for conditions which ensure a CLT. Notice that, due to the serial correlation inherent to the Markov chain, $\Sigma \neq \mathrm{Var}_F\{g(X)\}$ except in trivial cases. In Section 3 we propose two new estimators of Σ. For now, let Σ_n be a generic positive definite estimator of Σ.

A confidence region for μ constructed using Σ_n forms an ellipsoid in p dimensions oriented along the directions of the eigenvectors of Σ_n. Let $|\cdot|$ denote determinant. One can verify by straightforward calculation that the volume of the confidence region is proportional to $\sqrt{|\Sigma_n|}$ and thus depends on the estimated covariance matrix Σ_n only through the estimate $|\Sigma_n|$ of the *generalized variance of the Monte Carlo error*, $|\Sigma|$. The volume of the confidence region can describe whether the simulation effort is sufficiently large to achieve the desired level of precision in estimation [7, 19, 28].

Another common and intuitively reasonable method for choosing the simulation effort is to simulate until a desired effective sample size (ESS), i.e., the number with the property that μ_n has the same precision as the sample mean obtained by that number of independent and identically distributed (i.i.d.) samples, has been achieved [1, 5, 10]. Let $\Lambda = \mathrm{Var}_F\{g(X)\}$. Vats et al. [28] introduced the following definition of effective sample size

$$\mathrm{ESS} = n\left(\frac{|\Lambda|}{|\Sigma|}\right)^{\frac{1}{p}}\,, \tag{2}$$

which is naturally estimated with $n(|\Lambda_n|/|\Sigma_n|)^{1/p}$ where Λ_n is an estimator of Λ, e.g., the usual sample covariance matrix. Vats et al. [28] showed that terminating the simulation based on the effective sample size is equivalent to termination based on a relative confidence region where the Monte Carlo error is compared to size of the uncertainty in the target distribution. The point is that again a common method for assessing the reliability of the simulation is determined by the estimated generalized variance of the Monte Carlo error.

The estimators of Σ studied by Kosorok [21], Vats et al. [28], and Vats et al. [29] typically underestimate the generalized variance. We will propose a different method and show that it is asymptotically valid. Specifically, our method provides a consistent overestimate for the asymptotic generalized variance of the Monte Carlo error and therefore will result in a slightly larger simulation effort, leading to a more stable estimation process.

The rest of the paper is organized as follows. In Section 2 we develop notation and background in preparation for the estimation theory. In Section 3 we propose our method and establish its asymptotic validity. In Section 4 we examine the finite sample properties of the proposed method through a variety of examples. We consider a Bayesian logistic regression example of 5 covariates where a symmetric random walk Metropolis-Hastings algorithm is implemented to calculate the posterior mean of the regression coefficient vector, a Bayesian one-way random effects model where we use a random scan Gibbs sampler to estimate the posterior expectation of all 8 parameters, and a reversible multivariate AR(1) process that takes values in $\mathbb{R}^{12}$. We illustrate the use of multivariate methods in a meta-analysis application where the posterior has dimension 65.

2 Notation and Background

Recall that F has support $\mathcal{X}$ and let $\mathcal{B}(\mathcal{X})$ be a σ-algebra. For $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$ let $P^n(x, dy)$ be the n-step Markov transition kernel so that for $x \in \mathcal{X}$, $B \in \mathcal{B}(\mathcal{X})$, and $k \in \mathbb{N} = \{0, 1, 2, \ldots\}$ we have $P^n(x, B) = \Pr(X_{k+n} \in B | X_k = x)$, where $\Pr(\cdot)$ denotes probability. We assume that P satisfies detailed balance with respect to F. That is,

$$F(dx)P(x, dy) = F(dy)P(y, dx)\,. \tag{3}$$

Metropolis-Hastings algorithms satisfy (3) by construction as do many component-wise Markov chains, such as random scan or random sequence scan algorithms [16]. By integrating both sides of (3) it is easy to see that F is invariant for P. Suppose $X_0 \sim F$, that is the Markov chain is stationary. The assumption of stationarity is not crucial since, for Harris recurrent chains, if a CLT holds under stationarity, it holds for all initial distributions [23, Proposition 17.1.6].

The lag t autocovariance of the process $g(X_0), g(X_1), g(X_2), \ldots$ is defined as

$$\gamma_t = \gamma_{-t} = \mathrm{Cov}_F(g(X_i), g(X_{i+t}))\,.$$

Denote the sum of an adjacent pair of autocovariances by

$$\Gamma_i = \gamma_{2i} + \gamma_{2i+1}$$

for $i \in \mathbb{N}$ and its smallest eigenvalue by ξ_i.

We use the shorthand ∞ for $+\infty$ unless otherwise specified. If $\sum_{t=0}^{\infty} \gamma_t$ converges, the asymptotic covariance matrix in (1) can be written as [20]

$$\Sigma = \sum_{t=-\infty}^{+\infty} \gamma_t = -\gamma_0 + \sum_{t=0}^{\infty} (\gamma_t + \gamma_{-t}) = -\gamma_0 + 2\sum_{t=0}^{\infty} \gamma_t = -\gamma_0 + 2\sum_{i=0}^{\infty} \Gamma_i \,. \tag{4}$$

The following propositions will play a significant role in the development of the new estimation method in Section 3.

Proposition 1. *The following properties of the sequences* $\{\Gamma_i; i \in \mathbb{N}\}$ *and* $\{\xi_i; i \in \mathbb{N}\}$ *hold.*

1. Γ_i *is positive-definite, for all* $i \in \mathbb{N}$.
2. $\Gamma_i - \Gamma_{i+1}$ *is positive-definite, for all* $i \in \mathbb{N}$.
3. $\lim_{i\to\infty} \Gamma_i = 0$.
4. *The sequence* $\{\xi_i; i \in \mathbb{N}\}$ *is positive, decreasing, and converges to* 0.

Proof. See Appendix A. □

Recall (4) and let the mth partial sum be denoted

$$\Sigma_m = -\gamma_0 + \sum_{t=0}^{2m+1} (\gamma_t + \gamma_{-t}) = -\gamma_0 + 2\sum_{i=0}^{m} \Gamma_i \,. \tag{5}$$

Proposition 2. *The following properties of the sequence* $\{\Sigma_m; m \in \mathbb{N}\}$ *hold.*

1. *There exists a non-negative integer* m_0 *such that* Σ_m *is positive definite for* $m \geq m_0$ *and not positive definite for* $m < m_0$. *Specifically, when* $m_0 = 0$, Σ_m *is positive definite for all* m.
2. *The sequence* $\{|\Sigma_m|; m = m_0, m_0 + 1, m_0 + 2, \ldots\}$ *is positive, increasing, and converges to* $|\Sigma|$.

Proof. See Appendix A. □

Remark 1. The value of m_0 is difficult to calculate explicitly because Σ_m is usually not available in closed form. However, in Section 4.3 we consider a multivariate AR(1) Markov chain and verify that $m_0 = 0$. In the other examples, we cannot establish $m_0 = 0$ directly, but in our simulations we never observed anything else in 2000 independent replications.

3 Estimation Method

A natural estimator of the lagged autocovariance γ_t is the empirical autocovariance

$$\gamma_{n,t} = \gamma_{n,-t}^\top = \frac{1}{n}\sum_{i=1}^{n-t} \{g(X_i) - \mu_n\} \{g(X_{i+t}) - \mu_n\}^\top$$

where $\cdot^\top$ denotes transpose. Set

$$\widetilde{\gamma}_{n,t} = \frac{1}{2}(\gamma_{n,t} + \gamma_{n,-t})$$

for $t = 0, \ldots, n-1$ and write the sum of the ith ($0 \le i \le \lfloor n/2 - 1 \rfloor$) adjacent pair as

$$\Gamma_{n,i} = \widetilde{\gamma}_{n,2i} + \widetilde{\gamma}_{n,2i+1}\,.$$

By construction, $\Gamma_{n,i}$ is symmetric. Let $\xi_{n,i}$ denote its smallest eigenvalue. The empirical estimator of Σ_m ($0 \le m \le \lfloor n/2 - 1 \rfloor$) is

$$\Sigma_{n,m} = -\gamma_{n,0} + \sum_{t=0}^{2m+1} (\gamma_{n,t} + \gamma_{n,-t}) = -\gamma_{n,0} + 2\sum_{i=0}^{m} \Gamma_{n,i}. \tag{6}$$

Notice how (6) parallels (5).

3.1 Multivariate Initial Sequence Estimators

We are now in position to formally define the multivariate initial sequence (mIS) estimator. Let s_n be the smallest integer such that Σ_{n,s_n} is positive definite and let t_n be the largest integer m ($s_n \le m \le \lfloor n/2 - 1 \rfloor$) such that $|\Sigma_{n,i}| > |\Sigma_{n,i-1}|$ for all $i = s_n + 1, \ldots, m$. Then the mIS estimator, denoted $\Sigma_{\mathrm{seq},n}$, is defined as

$$\Sigma_{\mathrm{seq},n} = \Sigma_{n,t_n}\,.$$

It is possible that $\Sigma_{n,m}$ fails to be positive definite for all $m = 0, \ldots, \lfloor n/2 - 1 \rfloor$, and consequently s_n does not exist. Fortunately, when n is sufficiently large, we can always find such s_n.

Theorem 1. *With probability 1, s_n exists as $n \to \infty$. In particular, with probability 1, $s_n \to m_0$ as $n \to \infty$.*

Proof. See Appendix B. □

Thus mIS is feasible while the following establishes that it is asymptotically valid.

Theorem 2. *With probability 1,* $\liminf_{n\to\infty} |\Sigma_{\text{seq},n}| \geq |\Sigma|$.

Proof. See Appendix B. □

In the construction of $\Sigma_{\text{seq},n}$ we update $\Sigma_{n,i}$ to $\Sigma_{n,i+1} = \Sigma_{n,i} + 2\Gamma_{n,i+1}$. If $\Gamma_{n,i+1}$ has negative eigenvalues, adding $2\Gamma_{n,i+1}$ will squeeze the corresponding confidence region in undesirable directions. A remedy is to force the negative eigenvalues of $\Gamma_{n,i+1}$ to be 0. Suppose $\Gamma_{n,i+1}$ has eigen-decomposition $\Gamma_{n,i+1} = Q^\top \Lambda Q$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$. Define the positive part of $\Gamma_{n,i+1}$ as

$$\Gamma^+_{n,i+1} = Q^\top \Lambda^+ Q,$$

where $\Lambda^+ = \text{diag}(\max\{\lambda_1, 0\}, \ldots, \max\{\lambda_p, 0\})$.

This leads us to define the adjusted multivariate initial sequence (mISadj) estimator. Let s_n and t_n be as in the definition of mIS and let

$$\widetilde{\Sigma}_{n,t_n} = \Sigma_{n,s_n} + 2 \sum_{i=s_n+1}^{t_n} \Gamma^+_{n,i}$$

where $\Gamma^+_{n,i}$ is the positive part of $\Gamma_{n,i}$. Then the mISadj estimator, denoted $\Sigma_{\text{adj},n}$, is defined as

$$\Sigma_{\text{adj},n} = \widetilde{\Sigma}_{n,t_n}.$$

See Figure 1 for a display of the effect of using mISadj over mIS.

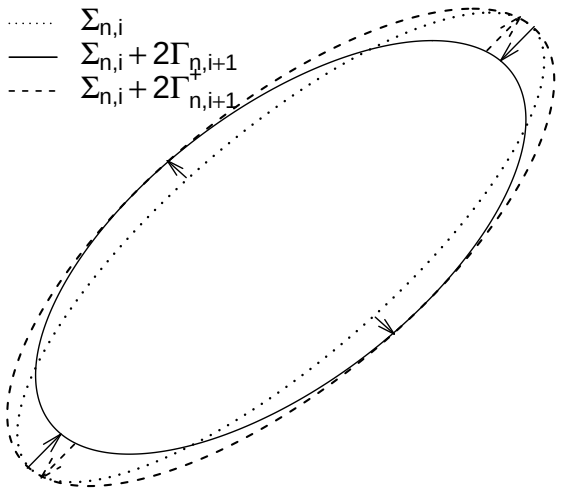

Figure 1: A diagrammatic sketch of the adjustment effect on a confidence region. Adding $2\Gamma_{n,i+1}$ squeezes the confidence region in the direction of the eigenvector corresponding to the negative eigenvalue. The adjustment cancels the shrinkage.

By construction, the mISadj estimator is positive definite. The modification adds a positive semi-definite matrix to the mIS estimator, which by Theorem 2 provides a consistent overestimate for the generalized variance, $|\Sigma|$, and therefore the mISadj estimator also has a larger determinant than the asymptotic covariance matrix, Σ.

Theorem 3. *With probability 1,* $\liminf_{n\to\infty} |\Sigma_{\text{adj},n}| \geq |\Sigma|$.

3.1.1 Related Estimators

The motivation for our approach can be found in Geyer's [9] univariate initial positive sequence (uIS) estimator. Suppose μ is one-dimensional and denote the variance of the asymptotic normal distribution σ^2. In this setting Geyer [9] proposed the uIS estimator:

$$\sigma^2_{\text{pos},n} = -\gamma_{n,0} + 2\sum_{i=0}^{t_n} \Gamma_{n,i}$$

where t_n is the largest integer m such that $\Gamma_{n,i} > 0$ for all $i = 1, \ldots, m$, That is, Geyer's truncation rule is to stop adding in $2\Gamma_{n,i}$ when it causes $\sigma^2_{n,i} = -\gamma_{n,0} + 2\sum_{j=0}^{i} \Gamma_{n,j}$ to decrease. (Figure 2 depicts the behavior of $\sigma^2_{n,i}$ and $\Gamma_{n,i}$ for one of the examples we consider later.) The uIS estimator is therefore the first local maximum of the sequence $\{\sigma^2_{n,i}; i = 0, \ldots, \lfloor n/2 - 1 \rfloor\}$ and thus gives an asymptotic overestimate of σ^2. This is formally stated in his Theorem 3.2:

$$\liminf_{n\to\infty} \sigma^2_{\text{pos},n} \geq \sigma^2 \quad \text{with probability 1.}$$

Neither mIS nor mISadj is a straightforward generalization of Geyer's method in that mIS and mISadj coincide but do not reduce to uIS when μ is one-dimensional. However, this is not essential because the three methods are asymptotically equivalent in univariate settings.

Kosorok [21] proposed an alternative multivariate estimator (mK) which was also motivated by Geyer's [9] approach. Recall from Proposition 1 that $\{\xi_i; i \in \mathbb{N}\}$ is positive, decreasing, and converges to 0, where ξ_i is the smallest eigenvalue of Γ_i. In mK the truncation point is chosen to be the largest integer m such that $\xi_{n,i} > 0$ for all $i = 1, \ldots, m$. However, this does not ensure that the generalized variance is adequately estimated and often truncates before the sequence $\{|\Sigma_{n,i}|; i = s_n, \ldots, \lfloor n/2 - 1 \rfloor\}$ reaches the first local maximum, as demonstrated in Figure 3 and 4.

4 Simulation Experiments

Our goal is to investigate the finite sample properties of mIS, mISadj, mK, and uIS through simulation experiments in a variety of examples. In each of the examples, which are described

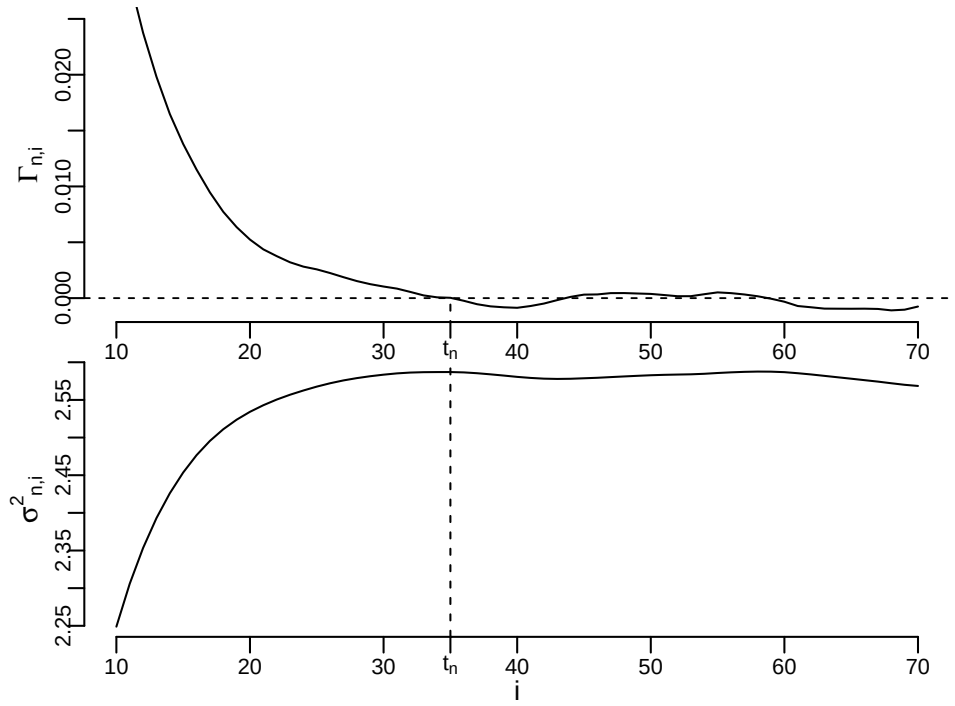

Figure 2: The uIS method truncates the first time $\Gamma_{n,i}$ is non-positive, or equivalently at the first local maximum of $\{\sigma^2_{n,i}; i = 0, \ldots, \lfloor n/2 - 1 \rfloor\}$. Computed using a marginal chain of the Bayesian logistic regression example described in Section 4.1 with Monte Carlo sample size 10^6.

in more detail below, we compare the approaches in terms of effective sample size as well as volume and coverage probability of a joint confidence region.

We describe the simulation examples and the MCMC algorithms used in Section 4.1–4.3. The results of the simulation experiments are given in Section 4.4. We then consider a meta-analysis application in Section 4.5.

4.1 Bayesian Logistic Regression

For $i = 1, \ldots, 100$, let $X_i = (x_{i1}, \ldots, x_{i5})$ be the observed covariates for the ith observation and Y_i be the binary response. We suppose

$$Y_i | X_i, \beta \overset{ind}{\sim} \text{Bernoulli}\left(\frac{1}{1 + \exp(-X_i\beta)}\right) \quad \text{and} \quad \beta \sim \mathcal{N}_5(0, 4I_5)\,.$$

This model results in a posterior on $\mathbb{R}^5$, denoted F. The data we use is provided in the `logit` dataset in the `mcmc` R package.

We are interested in estimating the posterior mean of β, that is, $\mu = \mathrm{E}_F\beta$. However,

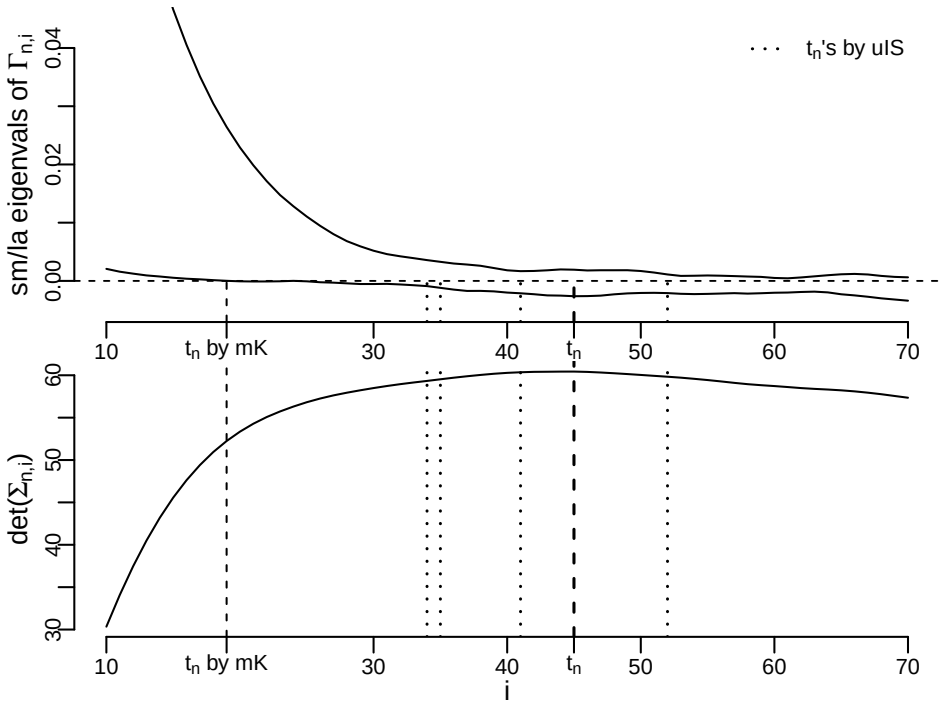

Figure 3: The thick dashed line marks the mIS truncation point. Truncating at the first local maximum of $\{|\Sigma_{n,i}|; i = s_n, \ldots, \lfloor n/2 - 1 \rfloor\}$ achieves a balance between the individual components. The mK truncation point–namely the first time $\{\Gamma_{n,i}; i = 0, \ldots, \lfloor n/2 - 1 \rfloor\}$ fails to be positive definite–is premature. Computed using the 5-dimensional Bayesian logistic regression example described in Section 4.1 with Monte Carlo sample size 10^6.

this expectation is intractable and hence we will use a symmetric random walk Metropolis-Hastings algorithm to estimate it. At each step of the Markov chain, the proposal for the next step is $\mathcal{N}_5(0,\, 0.3^2 I_5)$. The standard deviation of 0.3 ensures that in our application the acceptance rate is about 0.36.

By construction, the Metropolis-Hastings algorithm satisfies detailed balance (3). Vats et al. [28] established that this Markov chain is geometrically ergodic and that the posterior has a moment generating function and hence a CLT as at (1) holds.

4.2 Bayesian One-Way Random Effects Model

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

$$Y_i|\theta_i, \gamma_i \overset{ind.}{\sim} \mathcal{N}(\theta_i, \gamma_i^{-1}),$$
$$\theta_i|\mu, \lambda_\theta, \lambda_i \overset{ind.}{\sim} \mathcal{N}(\mu,\, \lambda_\theta^{-1}\lambda_i^{-1}),$$
$$\mu \sim \mathcal{N}(m_0, v_0^{-1}), \qquad \gamma_i \overset{i.i.d.}{\sim} \text{Gamma}(a_3, b_3),$$

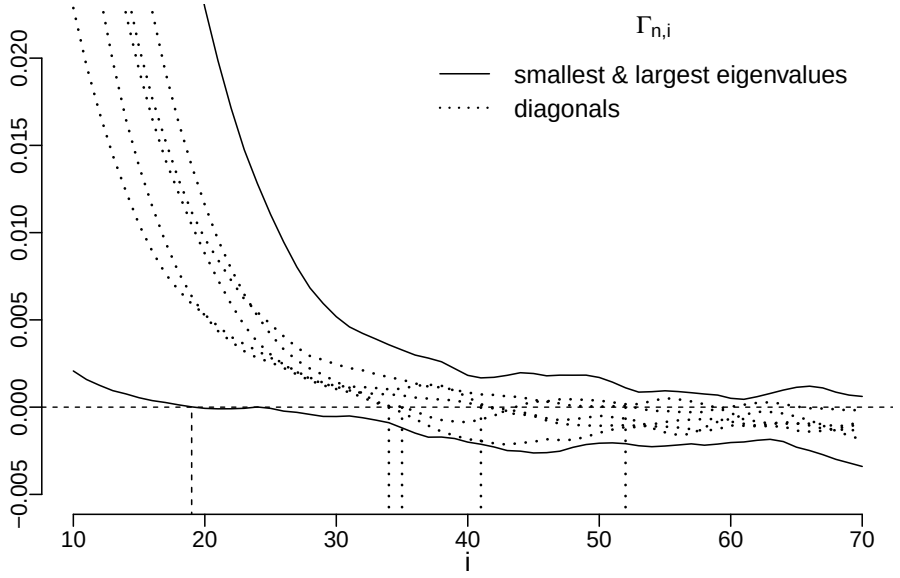

Figure 4: The diagonal entries of $\Gamma_{n,i}$, which correspond to the individual components, are always between the smallest and largest eigenvalues. It is too early to truncate the first time the smallest eigenvalue drops below 0, as is the case with mK (see the vertical dashed line). On the other extreme, it is too late to truncate when the largest eigenvalue drops below 0. The ideal truncation point should be somewhere between the uIS truncation points, marked by the vertical dotted lines. Computed using the 5-dimensional Bayesian logistic regression example described in Section 4.1 with Monte Carlo sample size 10^6.

$$\lambda_\theta \sim \text{Gamma}(a_1, b_1), \qquad \lambda_i \overset{i.i.d.}{\sim} \text{Gamma}(a_2, b_2),$$

where we assume the $a_1, a_2, a_3, b_1, b_2, b_3$ and v_0 are known positive constants while m_0 is a known scalar. We consider a data set simulated under the settings $K = 2$, $a_1 = a_2 = b_1 = b_2 = 0.1$, $a_3 = b_3 = 1.5$, $m_0 = 0$ and $v_0 = 0.001$. Let y denote all of the data, $\lambda = (\lambda_1, \ldots, \lambda_K)^\top$, $\xi = (\theta_1, \ldots, \theta_K, \mu)^\top$, and $\gamma = (\gamma_1, \ldots, \gamma_K)^\top$. The hierarchy results in a proper posterior density $f(\xi, \lambda_\theta, \lambda, \gamma|y)$ on $\mathbb{R}^{K+1} \times \mathbb{R}_+^{2K+1}$. One can verify that the posterior distribution has a finite second moment.

The posterior is intractable in the sense that posterior expectations are not generally available in closed form. We will use a random scan Gibbs sampler having the posterior as its invariant distribution to estimate the posterior expectation of all parameters. Doss and Hobert [4] derived the full conditional densities $f(\lambda_\theta|\xi, \lambda, \gamma)$, $f(\lambda|\xi, \lambda_\theta, \gamma)$, $f(\gamma|\xi, \lambda_\theta, \lambda)$, and $f(\xi|\lambda_\theta, \lambda, \gamma)$ required to implement random scan Gibbs.

It is well known that the random scan Gibbs sampler kernel is reversible, namely, satisfies detailed balance (3), with respect to the posterior (see e.g., Roberts and Rosenthal [26]).

Johnson and Jones [15] established geometric ergodicity of the random scan Gibbs sampler when $2a_1 + K - 2 > 0$ and $a_3 > 1$. These conditions combined with the second moment condition establish a Markov chain CLT.

4.3 Multivariate AR(1) Process

Consider an AR(1) process $\{X_n; n \in \mathbb{N}\}$ taking values in $\mathbb{R}^p$; that is,

$$X_{n+1} = AX_n + U_{n+1}$$

where U_n's are i.i.d. $\mathbb{R}^p$-valued random variables and A is a $p \times p$ matrix.

Ōsawa [24] proved that when U_n's follow a normal distribution $\mathcal{N}_p(\theta, V)$, then this $\mathbb{R}^p$-valued AR(1) process satisfies detailed balance (3) if and only if the matrix AV is symmetric. Suppose further that $\lim_{n\to\infty} A^n = 0$, then it has the stationary distribution $\mathcal{N}_p((I - A)^{-1}\theta, (I - A^2)^{-1}V)$. It is easy to verify that the second moment is finite.

Under stationarity one can derive the lag t autocovariance, $\gamma_t = A^{2t}(I - A^2)^{-1}V$, and hence the covariance matrix, $\Sigma = \{2(I - A^2)^{-1} - I\}(I - A^2)^{-1}V$, as in (4). Noticing that Σ is finite, and that the Markov chain is reversible with a finite second moment, we establish a Markov chain CLT (1) with mean $\mu = (I - A)^{-1}\theta$ and covariance matrix Σ [11, Corollary 6]. Also notice that $\Sigma_0 = \gamma_0 + 2\gamma_1$ is always positive definite, which satisfies the assumption in Remark 1 and hence guarantees the asymptotic properties of our proposed estimation method.

Let us consider the following choices that satisfy the conditions above: $\theta = \mathbf{1}_p$, $V = I_p$, and $A = p^{-1}H_p\mathrm{diag}(2^{-1}, \ldots, 2^{-p})H_p^\top$, where H_p is a Hadamard matrix of order p. We set $p = 12$ in our simulation study.

4.4 Results

In this section we refer to the setting of Section 4.1 as example 1, the setting of Section 4.2 as example 2, and the setting of Section 4.3 as example 3. For all examples we ran 2000 independent replications of the Markov chain for 10^6 iterations in example 1 and 3 and 5×10^5 iterations in example 2, respectively. We will compare the multivariate methods–namely mIS, mISadj, and mK–in the context of estimating the effective sample size. We then turn our attention to the finite sample properties of the confidence regions produced by the multivariate methods, yielding ellipsoidal regions, and Geyer's univariate uIS for individual components, yielding cube-shaped regions. To assess coverage probabilities in example 1 and 2 we perform an independent run of length 10^{10} of the Markov chain in each example and declared the sample average over those 10^{10} iterations to be the truth, while in example 3, the true mean is obtained through the closed form expression derived.

Table 1: Estimated ESS with standard errors. For uIS, only the minimum estimated ESS is reported.

	mK	mIS	mISadj	uIS
Ex1($\times 10^4$)	5.40 (.002)	5.22 (.001)	5.18 (.001)	3.95 (.002)
Ex2($\times 10^4$)	4.74 (.007)	3.76 (.002)	3.52 (.003)	1.30 (.001)
Ex3($\times 10^5$)	8.78 (.000)	8.39 (.000)	8.30 (.001)	7.58 (.001)

The results concerning effective sample size of the simulation experiments are given in Table 1. Prior to the work of Vats et al. [28] it was standard to report the minimum of the univariate effective sample size calculated component-wise. This leads to a substantial underestimate of the effective sample size as can be seen in Table 1. In contrast, multivariate error estimation yields more accurate evaluation of the effective sample size. We can approximately order the multivariate methods in terms of estimated effective sample size: mK $>$ mIS $>$ mISadj. That is, mK is more optimistic than mIS and mISadj.

We construct 90% confidence regions using the multivariate estimation methods and uIS. Throughout "uIS" and "uIS-Bonferroni" represent the uncorrected and Bonferroni corrected confidence regions generated by uIS, respectively. Let us first examine the volumes of the confidence regions generated by different methods.

Table 2: Average volumes to the pth ($p = 5, 8, 12$ for Ex1, 2, 3) root and standard errors of nominal 90% confidence regions.

	uIS	mK	mIS	mISadj	uIS-Bonferroni
Ex1($\times 10^{-3}$)	5.53 (.001)	6.31 (.001)	6.41 (.001)	6.44 (.001)	7.82 (.001)
Ex2($\times 10^{-2}$)	3.51 (.002)	3.95 (.003)	4.43 (.003)	4.58 (.003)	5.33 (.004)
Ex3($\times 10^{-3}$)	3.84 (.000)	4.78 (.000)	4.89 (.000)	4.92 (.000)	6.16 (.000)

The volumes are presented in ascending order from left to right across Table 2. The uncorrected uIS confidence regions are much smaller than the other methods, while the Bonferroni correction considerably enlarges the confidence regions, resulting in bigger volumes than all the multivariate methods.

Recall that the volume of a confidence region depends on the estimated covariance matrix only through the estimated generalized variance of the Monte Carlo error. Therefore, Table 2 compares the estimation of the generalized variance by different multivariate methods. We observe that mK underestimates the generalized variance relatively to mIS. The mISadj

method is comparable to mIS in example 1 and 3 but clearly overestimates in example 2.

Table 3: Estimated coverage probabilities and standard errors of nominal 90% confidence regions.

	uIS	mK	mIS	mISadj	uIS-Bonferroni
Ex1	.622 (.0108)	.885 (.0071)	.898 (.0068)	.900 (.0067)	.908 (.0065)
Ex2	.386 (.0109)	.660 (.0106)	.845 (.0081)	.881 (.0073)	.862 (.0077)
Ex3	.323 (.0105)	.882 (.0072)	.911 (.0064)	.916 (.0062)	.917 (.0062)

The proposed method, mIS, exceeds mK in both the volume and the coverage of confidence regions, although the coverage rate does not always reach the expectation. The adjustment moderately increases the coverage probability.

The uncorrected uIS regions have a poor coverage. The Bonferroni regions work well in these examples, but in high-dimensional cases the Bonferroni correction can be overly conservative. Overall, multivariate error estimation methods yield better confidence regions.

4.5 A Meta-Analysis Example

Doss and Hobert [4] carried out meta-analyses to study the effect of non-steroidal anti-inflammatory drugs (NSAIDs) on the risk of colon cancer. The dataset consists of 21 studies that relate NSAIDs intake and risk of colon cancer (see Harris et al. [12] and Doss and Hobert [4] for details). We apply the Bayesian one-way random effects model described in Section 4.2 to the colon cancer dataset. The posterior $f(\theta_1, \ldots, \theta_K, \mu, \lambda_\theta, \lambda_1, \ldots, \lambda_K, \gamma_1, \ldots, \gamma_K | y)$ has dimension $p = 65$ when $K = 21$.

We run a Markov chain for 4×10^6 iterations and compute the multivariate estimators–namely mIS, mISadj, and mK–along with Geyer's uIS for individual components.

Table 4 shows the estimated effective sample sizes. The uIS method results in 65 estimated effective sample sizes, each of which corresponds to a component of the posterior distribution. Only the minimum estimated univariate effective sample size is reported.

Table 4: Estimated ESS ($\times 10^5$) with Monte Carlo sample size 4×10^6 using the colon cancer dataset. For uIS, only the minimum estimated ESS is reported.

mK	mIS	mISadj	uIS
4.637	4.296	4.134	1.137

An advantage of using multivariate methods like mIS over univariate estimation like uIS is that only multivariate methods capture the cross-correlation between components. This cross-correlation is often significant as seen in Figure 5.

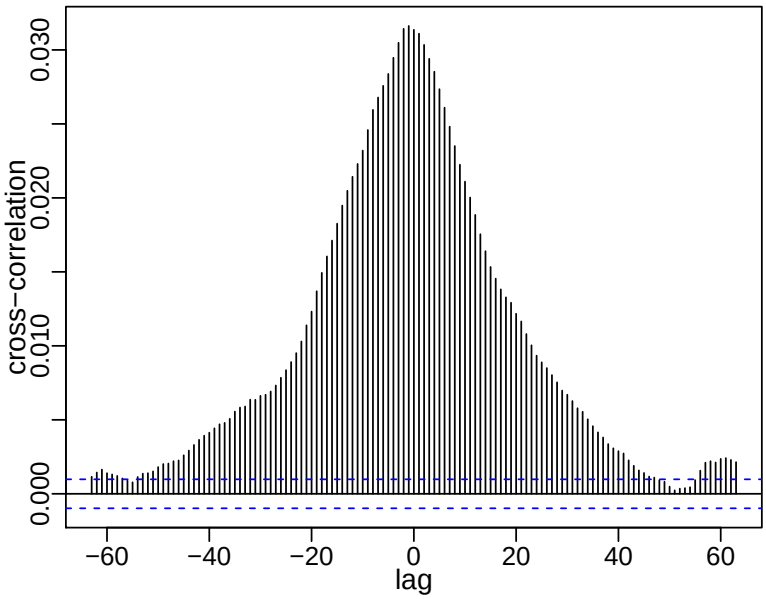

Figure 5: Cross-correlation plot between μ and λ_1. Computed with Monte Carlo sample size 4×10^6 using the colon cancer dataset.

We construct 90% confidence regions using the multivariate estimation methods and uIS. The left panel of Figure 6 shows the cross-sections of the confidence regions that are cut through the center of the confidence regions parallel to the plane spanned by μ and λ_1. The reader should not be worried that the cross-sectioned ellipsoids appear much larger than the Bonferroni region. The full 65-dimensional ellipsoid will have a smaller volume than the 65-dimensional Bonferroni region, but this does not have to be the case for cross-sectioned regions. As a comparison, in the right panel of Figure 6 we present bivariate 90% confidence regions for μ and λ_1 when we ignore the other 63 components. This clearly shows how multivariate estimation methods generate confidence regions that are not so liberal as uIS, yet not so conservative as uIS-Bonferroni.

Table 5 compares the volumes of the confidence regions generated by different methods. The results agree with our conclusion from the previous simulation study: mISadj is slightly more conservative than mIS; mK clearly underestimates the generalized variance. The volumes generated by multivariate estimators are fairly close to each other but the univariate results are far away. Apparently uIS is too liberal while uIS-Bonferroni is too conservative, but the multivariate methods achieve a balance.

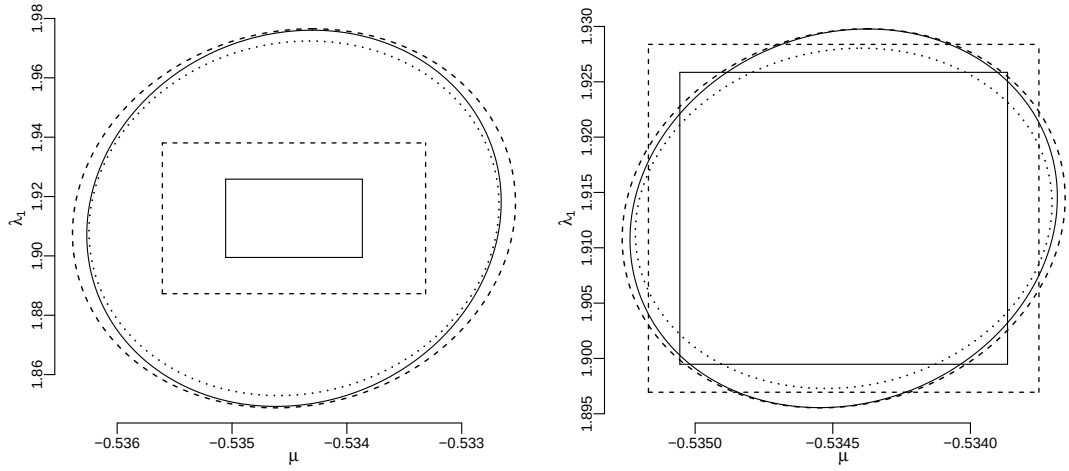

Figure 6: 90% confidence regions for (μ, λ_1) computed with Monte Carlo sample size 4×10^6 using the colon cancer dataset. The left panel displays confidence regions based on a cross-section of the $p = 65$-dimensional region parallel to the plane spanned by (μ, λ_1). In the right panel the confidence regions are created by ignoring the other 63 components. In both panels, the solid ellipsoid, dashed ellipsoid, and dotted ellipsoid corresponds to mIS, mISadj, and mK, respectively while the small solid and the big dashed rectangles are uIS and uIS-Bonferroni, respectively.

Table 5: Volumes to the pth root ($\times 10^{-3}$) of 90% confidence regions for all components ($p = 65$) and for μ and λ_1 only ($p = 2$). Computed with Monte Carlo sample size 4×10^6 using the colon cancer dataset.

uIS	mK	mIS	mISadj	uIS-Bonferroni
6.96	8.70	9.04	9.22	13.39

(a) 65-dimensional confidence regions.

uIS	mK	mIS	mISadj	uIS-Bonferroni
5.60	6.04	6.44	6.56	6.68

(b) Bivariate confidence regions for μ and λ_1.

4.6 Discussion

The preceding simulation examples and the theory developed indicate that mIS and mISadj perform as they were designed to in that they provide a consistent overestimate of the asymptotic generalized variance of the Monte Carlo error. Compared to standard univariate

methods, our estimators adjust for multivariate issues and thus provide more realistic estimates of Monte Carlo effective sample size and slightly larger confidence regions which result in improved performance in terms of coverage probabilities.

Acknowledgments

The authors are grateful to Charles Geyer and Dootika Vats for helpful conversations.

Appendices

A Proofs of Propositions 1 and 2

We begin with some preliminary results which will be useful later.

Lemma 1 (Harville [13], Lemma 18.2.17)**.** *Let $A_0, A_1, A_2, \ldots$ represent a sequence of $m \times n$ matrices. If the infinite series $\sum_{k=0}^{\infty} A_k$ converges, then $\lim_{k\to\infty} A_k = 0$.*

Since the eigenvalues of a Hermitian $p \times p$ matrix A are real, we may (and do) adopt the convention that they are always arranged in algebraically non-decreasing order:

$$\lambda_{\min}(A) = \lambda^{(1)}(A) \leq \lambda^{(2)}(A) \leq \ldots \leq \lambda^{(p-1)}(A) \leq \lambda^{(p)}(A) = \lambda_{\max}(A)\,. \tag{7}$$

Lemma 2 (Horn and Johnson [14], Corollary 4.3.15)**.** *Let $p \times p$ matrices A, B be Hermitian and let the respective eigenvalues of A, B, and $A+B$ be $\{\lambda^{(k)}(A); k = 1, \ldots, p\}$, $\{\lambda^{(k)}(B); k = 1, \ldots, p\}$, and $\{\lambda^{(k)}(A+B); k = 1, \ldots, p\}$, each algebraically ordered as in* (7)*. Then*

$$\lambda^{(k)}(A) + \lambda^{(1)}(B) \leq \lambda^{(k)}(A+B), \qquad k = 1, \ldots, p. \tag{8}$$

Lemma 3. *Suppose we have two $p \times p$ Hermitian matrices A and B. Let the respective eigenvalues of A and B be $\{\lambda^{(k)}(A); k = 1, \ldots, p\}$ and $\{\lambda^{(k)}(B); k = 1, \ldots, p\}$, each algebraically ordered as in* (7)*. If $A - B$ is positive definite, then $\lambda^{(k)}(A) > \lambda^{(k)}(B)$, for all $k = 1, \ldots, p$. Further, if A and B are both positive semi-definite, then $|A| > |B|$.*

Proof. Applying (8) to B and $A - B$, we get for all $k = 1, \ldots, p$,

$$\lambda^{(k)}(B) + \lambda^{(1)}(A - B) \leq \lambda^{(k)}(A).$$

Since $A - B$ is positive definite, we have

$$\lambda^{(1)}(A - B) > 0.$$

Therefore, for all $k = 1, \ldots, p$,

$$\lambda^{(k)}(A) > \lambda^{(k)}(B).$$

When A and B are both positive semi-definite, we further have for all $k = 1, \ldots, p$,

$$\lambda^{(k)}(A) > \lambda^{(k)}(B) \geq 0. \tag{9}$$

Since the determinant is equal to the product of all eigenvalues, we take product of (9) for all k and obtain

$$|A| > |B|.$$

□

Lemma 4 (Vats et al. [29], Theorem 2)**.** *Let Σ_n be a strongly consistent estimator of Σ. Let the respective eigenvalues of Σ_n and Σ be $\{\lambda^{(k)}(\Sigma_n); k = 1, \ldots, p\}$ and $\{\lambda^{(k)}(\Sigma); k = 1, \ldots, p\}$, each algebraically ordered as in (7). Then $\lambda^{(k)}(\Sigma_n) \to \lambda^{(k)}(\Sigma)$ with probability 1 as $n \to \infty$ for all $k = 1, \ldots, p$.*

Corollary 1. *Let Σ_n be a strongly consistent estimator of Σ, then $|\Sigma_n| \to |\Sigma|$ with probability 1 as $n \to \infty$.*

A.1 Proof of Proposition 1

We begin with the univariate case so $g : \mathcal{X} \to \mathbb{R}$. Let $\mathcal{E}$ be the spectral decomposition measure associated with transition kernel P and $\mathcal{E}_g$ be the induced spectral measure for g. Details on the spectral decomposition measure can be found in Rudin [27], Chan and Geyer [2], and Häggström and Rosenthal [11]. Specifically, for all $t \in \mathbb{N}$,

$$\gamma_t = \int_{-1}^{1} \lambda^t \mathcal{E}_g(d\lambda)\,. \tag{10}$$

It follows that for all $i \in \mathbb{N}$,

$$\Gamma_i = \gamma_{2i} + \gamma_{2i+1} = \int_{-1}^{1} \lambda^{2i}(1+\lambda)\mathcal{E}_g(d\lambda)$$

and

$$\Gamma_i - \Gamma_{i+1} = \int_{-1}^{1} \lambda^{2i}(1+\lambda)^2(1-\lambda)\mathcal{E}_g(d\lambda)\,.$$

Therefore, Γ_i and $\Gamma_i - \Gamma_{i+1}$ must be non-negative. To prove Proposition 1(1) that $\Gamma_i > 0$ and Proposition 1(2) that $\Gamma_i - \Gamma_{i+1} > 0$, we need to show that neither Γ_i nor $\Gamma_i - \Gamma_{i+1}$ can be zero.

For $i = 0$,

$$\Gamma_0 = \int_{-1}^{1} (1+\lambda)\mathcal{E}_g(d\lambda) = 0 \text{ if and only if } \mathcal{E}_g(\{-1\}) = 1, \tag{11}$$

and

$$\Gamma_0 - \Gamma_1 = \int_{-1}^{1} (1+\lambda)^2(1-\lambda)\mathcal{E}_g(d\lambda) = 0 \text{ if and only if } \mathcal{E}_g(\{-1,1\}) = 1\,. \tag{12}$$

For $i \in \mathbb{N}^+$,

$$\Gamma_i = \int_{-1}^{1} \lambda^{2i}(1+\lambda)\mathcal{E}_g(d\lambda) = 0 \text{ if and only if } \mathcal{E}_g(\{-1,0\}) = 1, \tag{13}$$

and

$$\Gamma_i - \Gamma_{i+1} = \int_{-1}^{1} \lambda^{2i}(1+\lambda)^2(1-\lambda)\mathcal{E}_g(d\lambda) = 0 \text{ if and only if } \mathcal{E}_g(\{-1,0,1\}) = 1\,. \tag{14}$$

By (11)–(14), for an arbitrary $i \in \mathbb{N}$, a necessary condition for each of $\Gamma_i = 0$ and $\Gamma_i - \Gamma_{i+1} = 0$ is $\mathcal{E}_g(\{-1,0,1\}) = 1$. We now show that $\mathcal{E}_g(\{-1,0,1\}) = 1$ cannot hold under our assumptions, so that both Γ_i and $\Gamma_i - \Gamma_{i+1}$ are non-zero, which completes the proof of Proposition 1(1)(2).

If $\mathcal{E}_g$ is a point mass at 0, then (10) yields

$$\gamma_t = \int_{-1}^{1} \lambda^t \mathcal{E}_g(d\lambda) = 0$$

for all $t \in \mathbb{N}^+$, which is trivial. Therefore, without loss of generality, we assume

$$\mathcal{E}_g(\{0\}) < 1. \tag{15}$$

Häggström and Rosenthal [11] showed that when P is irreducible and aperiodic,

$$\mathcal{E}_g(\{-1,1\}) = 0\,. \tag{16}$$

It follows from (15) and (16) that $\mathcal{E}_g(\{-1,0,1\}) < 1$. By previous arguments, we have proved Proposition 1(1)(2). That is, for all $i \in \mathbb{N}$, $\Gamma_i > 0$ and $\Gamma_i - \Gamma_{i+1} > 0$.

Proposition 1(3), namely $\lim_{i\to\infty} \Gamma_i = 0$, follows from Lemma 1 and the assumption that $\sum_{i=0}^{\infty} \Gamma_i$ converges.

Finally, by Proposition 1(1)(2)(3), we obtain Proposition 1(4), that is, $\{\Gamma_i; i \in \mathbb{N}\}$ is positive, decreasing, and converges to 0.

We now turn to the multivariate case so $g : \mathcal{X} \to \mathbb{R}^p$ and $p \geq 2$. Set $h = v^\top g$ for an arbitrary $v \in \mathbb{R}^p$ and $v \neq 0$. Then $h : \mathcal{X} \to \mathbb{R}$ is measurable and square integrable with respect to F. Recall that the Markov chain is assumed stationary. For $t \in \mathbb{N}$ define the lag t autocovariance

$$\gamma_t^* = \gamma_{-t}^* = \mathrm{Cov}_F(h(X_i), h(X_{i+t}))$$

and for $i \in \mathbb{N}$ define

$$\Gamma_i^* = \gamma_{2i}^* + \gamma_{2i+1}^*\,.$$

Notice that

$$\begin{aligned}\Gamma_i^* &= \gamma_{2i}^* + \gamma_{2i+1}^* \\ &= \mathrm{Cov}_F(h(X_0), h(X_{2i})) + \mathrm{Cov}_F(h(X_0), h(X_{2i+1})) \\ &= v^\top \mathrm{Cov}_F(g(X_0), g(X_{2i}))v + v^\top \mathrm{Cov}_F(g(X_0), g(X_{2i+1}))v \\ &= v^\top \gamma_{2i} v + v^\top \gamma_{2i+1} v \\ &= v^\top (\gamma_{2i} + \gamma_{2i+1}) v \\ &= v^\top \Gamma_i v\,.\end{aligned}$$

By the univariate case considered above, $\Gamma_i^* > 0$. Since v is arbitrary, Γ_i is positive definite. A similar argument shows that $\Gamma_i - \Gamma_{i+1}$ is positive definite. This establishes Proposition 1(1)(2).

Use Lemma 1 and notice that $\sum_{i=0}^{\infty} \Gamma_i$ converges by assumption. We obtain $\lim_{i\to\infty} \Gamma_i = 0$. Thus Proposition 1(3) is proved.

Since Γ_i is positive definite, $\xi_i > 0$ for all $i \in \mathbb{N}$. Since $\Gamma_i - \Gamma_{i+1}$ is positive definite we obtain from Lemma 3 that $\xi_i > \xi_{i+1}$. Hence $\xi_i \to 0$ as $i \to \infty$ which establishes Proposition 1(4).

A.2 Proposition 2

For all $m \in \mathbb{N}$ let λ_m be the smallest eigenvalue of Σ_m. Notice that $\Sigma_m - \Sigma_{m-1} = 2\Gamma_m$ is positive definite by Proposition 1. Then Lemma 3 implies $\lambda_m > \lambda_{m-1}$ and hence$\{\lambda_m; m \in \mathbb{N}\}$ is monotonically increasing. Since $\{\Sigma_m; m \in \mathbb{N}\}$ converges to the asymptotic covariance matrix Σ, by Lemma 4 we have

$$\lim_{m\to\infty} \lambda_m = \lambda > 0,$$

where λ is the smallest eigenvalue of Σ.

If $\lambda_0 \le 0$, there exists a positive integer m_0 such that $\lambda_m > 0$ for $m \ge m_0$ and $\lambda_m \le 0$ for $m < m_0$. If $\lambda_0 > 0$, then $\lambda_m > 0$ for all $m \in \mathbb{N}$. In this case, let $m_0 = 0$. Immediately we have that Σ_m is positive definite for $m \ge m_0$ and not positive definite for $m < m_0$. It then follows that for all $m \ge m_0$, $|\Sigma_m| > 0$.

Now let $m > m_0$ and notice that $\Sigma_m - \Sigma_{m-1}$ is positive definite. Using Lemma 3 we obtain for all $m > m_0$,

$$|\Sigma_m| > |\Sigma_{m-1}|\,.$$

By $\lim_{m\to\infty} \Sigma_m = \Sigma$ and Corollary 1,

$$\lim_{m\to\infty} |\Sigma_m| = |\Sigma|\,.$$

Therefore, $\{|\Sigma_m|; m = m_0, m_0+1, m_0+2, \ldots\}$ is positive, increasing, and converges to $|\Sigma|$.

B Proofs of Theorems 1 and 2

Lemma 5. *For all $t \in \mathbb{N}$, with probability 1, as $n \to \infty$,*

$$\gamma_{n,t} \to \gamma_t.$$

Proof. Notice that

$$\begin{aligned}\gamma_{n,t} =& \frac{1}{n}\sum_{i=1}^{n-t} \{g(X_{n,i}) - \mu_n\} \{g(X_{n,i+t}) - \mu_n\}^\top \\ =& \frac{1}{n}\sum_{i=1}^{n-t} g(X_{n,i}) g(X_{n,i+t})^\top - \frac{1}{n}\sum_{i=1}^{n-t} g(X_{n,i})\mu_n^\top - \frac{1}{n}\mu_n \sum_{i=1}^{n-t} g(X_{n,i+t})^\top + \frac{n-t}{n}\mu_n\mu_n^\top \,.\end{aligned}$$

By repeated application of the Markov chain strong law we see that, with probability 1, as $n \to \infty$,

$$\gamma_{n,t} \to \mathrm{E}_F\left\{g(X_0)g(X_t)^\top\right\} - \mu\mu^\top = \mathrm{Cov}_F\left(g(X_0), g(X_t)\right) = \gamma_t\,.$$

□

Corollary 2. *For all $m \in \mathbb{N}$, with probability 1, as $n \to \infty$,*

$$\Sigma_{n,m} \to \Sigma_m\,.$$

Proof. This follows immediately from Lemma 5. □

Lemma 6. *If a sequence of random variables $X_1, X_2, \ldots$ converges to X with probability 1, then, for an arbitrary $x \in \mathbb{R}$ such that $\Pr(X = x) = 0$,*

$$\liminf_{n\to\infty}\{X_n \le x\} = \{X \le x\} \text{ w.p. } 1 \tag{17}$$

and

$$\liminf_{n\to\infty}\{X_n > x\} = \{X > x\} \text{ w.p. } 1. \tag{18}$$

Proof. We only prove the first part. The second part can be shown by a similar argument.

Recall that two events A and B are equal almost surely if both of the events $A\backslash B$ and $B\backslash A$ are null sets [8, p. 13]. Thus we need only show that both $\liminf_{n\to\infty}\{X_n \le x\} \setminus \{X \le x\}$ and $\{X \le x\} \setminus \liminf_{n\to\infty}\{X_n \le x\}$ are null sets.

Suppose $\omega \in \liminf_{n\to\infty}\{X_n \le x\} \setminus \{X \le x\}$. By definition,

$$\omega \in \liminf_{n\to\infty}\{X_n \le x\}$$

is equivalent to saying that there exists some n such that for all $m \geq n$, $X_m(\omega) \leq x$. This implies that

$$\lim_{n\to\infty} X_n(\omega) \leq x < X(\omega),$$

where the second inequality is due to $\omega \notin \{X \leq x\}$. It follows that

$$\omega \in \left\{ \lim_{n\to\infty} X_n \neq X \right\}.$$

Thus we have that

$$\liminf_{n\to\infty} \{X_n \leq x\} \setminus \{X \leq x\} \subset \left\{ \lim_{n\to\infty} X_n \neq X \right\}$$

which is a null set because $X_n \overset{a.s.}{\to} X$.

Suppose $\omega \in \{X \leq x\} \setminus \liminf_{n\to\infty}\{X_n \leq x\}$. By definition,

$$\omega \notin \liminf_{n\to\infty} \{X_n \leq x\}$$

is equivalent to saying that for all n, there exists some $m \geq n$ such that $X_m(\omega) > x$. This implies that

$$\lim_{n\to\infty} X_n(\omega) \geq x \geq X(\omega),$$

where the second inequality is due to $\omega \in \{X \leq x\}$. It follows that

$$\omega \in \left\{ \lim_{n\to\infty} X_n \neq X \right\} \bigcup \{X = x\}.$$

Thus we have that

$$\{X \leq x\} \setminus \limsup_{n\to\infty} \{X_n \leq x\} \subset \left\{ \lim_{n\to\infty} X_n \neq X \right\} \bigcup \{X = x\},$$

which is a null set.

So far we have proved (17). A similar argument can be used to prove (18). □

Equipped with the preceding results, we now prove the following lemma in preparation for Theorems 1 and 2.

Recall that m_0 is a non-negative integer such that Σ_m is positive definite for $m \geq m_0$ and not positive definite for $m < m_0$. Also recall that s_n is the smallest integer such that Σ_{n,s_n} is positive definite and that t_n is the largest integer m ($s_n \leq m \leq \lfloor n/2 - 1 \rfloor$) such that $|\Sigma_{n,i}| > |\Sigma_{n,i-1}|$ for all $i = s_n + 1, \ldots, m$. The smallest eigenvalues of Σ_m and $\Sigma_{n,m}$ are denoted λ_m and $\lambda_{n,m}$, respectively.

Lemma 7. *Suppose* $\liminf_{n\to\infty}\{\lambda_{n,m_0-1} \leq 0\}$ *occurs with probability 1. For all* $K \geq m_0$,

$$\Pr\left(\liminf_{n\to\infty}\{s_n = m_0, t_n \geq K\}\right) = 1.$$

Proof. Define $\Delta_m = |\Sigma_m| - |\Sigma_{m-1}|$ and $\Delta_{n,m} = |\Sigma_{n,m}| - |\Sigma_{n,m-1}|$. Notice that

$$\begin{aligned}
&\{s_n = m_0, t_n \geq K\} \\
&= \{s_n = m_0\} \bigcap \{\Delta_{n,i} > 0 \text{ for all } i \text{ such that } m_0 < i \leq K\} \\
&= \left(\bigcap_{m<m_0} \{\Sigma_{n,m} \text{ is not positive definite}\} \right) \bigcap \{\Sigma_{n,m_0} \text{ is positive definite}\} \\
&\quad \bigcap \left(\bigcap_{m_0<i\leq K} \{\Delta_{n,i} > 0\} \right) \\
&= \left(\bigcap_{m<m_0} \{\lambda_{n,m} \leq 0\} \right) \bigcap \{\lambda_{n,m_0} > 0\} \bigcap \left(\bigcap_{m_0<i\leq K} \{\Delta_{n,i} > 0\} \right),
\end{aligned}$$

where $\lambda_{n,m}$ denotes the smallest eigenvalue of $\Sigma_{n,m}$. Then we write

$$\begin{aligned}
&\liminf_{n\to\infty} \{s_n = m_0, t_n \geq K\} \\
&= \left(\bigcap_{m<m_0} \liminf_{n\to\infty} \{\lambda_{n,m} \leq 0\} \right) \bigcap \liminf_{n\to\infty} \{\lambda_{n,m_0} > 0\} \bigcap \left(\bigcap_{m_0<i\leq K} \liminf_{n\to\infty} \{\Delta_{n,i} > 0\} \right). \quad (19)
\end{aligned}$$

By Lemma 4, Corollary 1 and Corollary 2, for all m, with probability 1,

$$\lambda_{n,m} \to \lambda_m \text{ as } n \to \infty, \quad (20)$$

and for all i, with probability 1,

$$\Delta_{n,i} \to \Delta_i \text{ as } n \to \infty. \quad (21)$$

By Proposition 2(1), $\lambda_{m_0} > 0$ so that $\Pr(\lambda_{m_0} = 0) = 0$ and $\lambda_m \leq 0$ for all $m < m_0$. In particular, $\lambda_m < 0$ so that $\Pr(\lambda_m = 0) = 0$ for all $m < m_0 - 1$. By Proposition 2(2), $\Delta_i > 0$ so that $\Pr(\Delta_i = 0) = 0$ for $i > m_0$. Then by Lemma 6 we have that for all $m < m_0 - 1$,

$$\liminf_{n\to\infty} \{\lambda_{n,m} \leq 0\} \overset{a.s.}{=} \left\{ \lim_{n\to\infty} \lambda_{n,m} \leq 0 \right\}, \quad (22)$$

$$\liminf_{n\to\infty} \{\lambda_{n,m_0} > 0\} \overset{a.s.}{=} \left\{ \lim_{n\to\infty} \lambda_{n,m_0} > 0 \right\},$$

and for $i > m_0$

$$\liminf_{n\to\infty} \{\Delta_{n,i} > 0\} \overset{a.s.}{=} \left\{ \lim_{n\to\infty} \Delta_{n,i} > 0 \right\}.$$

Notice that (22) holds for $m = m_0 - 1$ if $\lambda_{m_0-1} < 0$. When $\lambda_{m_0-1} = 0$, (22) is true only if $\liminf_{n\to\infty} \{\lambda_{n,m_0-1} \leq 0\}$ occurs with probability 1.

Under the preceding assumption, we continue to write (19) as

$$\liminf_{n\to\infty} \{s_n = m_0, t_n \geq K\}$$

$$= \left(\bigcap_{m<m_0} \liminf_{n\to\infty} \{\lambda_{n,m} \leq 0\} \right) \bigcap \liminf_{n\to\infty} \{\lambda_{n,m_0} > 0\} \bigcap \left(\bigcap_{m_0<i\leq K} \liminf_{n\to\infty} \{\Delta_{n,i} > 0\} \right)$$

$$\overset{a.s.}{=} \left(\bigcap_{m<m_0} \left\{ \lim_{n\to\infty} \lambda_{n,m} \leq 0 \right\} \right) \bigcap \left\{ \lim_{n\to\infty} \lambda_{n,m_0} > 0 \right\} \bigcap \left(\bigcap_{m_0<i\leq K} \left\{ \lim_{n\to\infty} \Delta_{n,i} > 0 \right\} \right). \quad (23)$$

By Proposition 2(1), $\lambda_{m_0} > 0$ and $\lambda_m \leq 0$ for all $m < m_0$. Then by (20) we have for $m < m_0$

$$\Pr\left(\lim_{n\to\infty} \lambda_{n,m} \leq 0 \right) \geq \Pr\left(\lim_{n\to\infty} \lambda_{n,m} = \lambda_m \right) = 1, \quad (24)$$

and

$$\Pr\left(\lim_{n\to\infty} \lambda_{n,m_0} > 0 \right) \geq \Pr\left(\lim_{n\to\infty} \lambda_{n,m_0} = \lambda_{m_0} \right) = 1. \quad (25)$$

By Proposition 2(2), $\Delta_i > 0$ for $i > m_0$. Then by (21) we have

$$\Pr\left(\lim_{n\to\infty} \Delta_{n,i} > 0 \right) \geq \Pr\left(\lim_{n\to\infty} \Delta_{n,i} = \Delta_i \right) = 1. \quad (26)$$

It follows from (24)-(26) that

$$\Pr\left\{ \left(\bigcap_{m<m_0} \left\{ \lim_{n\to\infty} \lambda_{n,m} \leq 0 \right\} \right) \bigcap \left\{ \lim_{n\to\infty} \lambda_{n,m_0} > 0 \right\} \bigcap \left(\bigcap_{m_0<i\leq K} \left\{ \lim_{n\to\infty} \Delta_{n,i} > 0 \right\} \right) \right\} = 1.$$

Then by (23) we obtain the result. □

Remark 2. Consider the assumption that $\liminf_{n\to\infty}\{\lambda_{n,m_0-1} \leq 0\}$ occurs with probability 1. If $m_0 = 0$, then this assumption is not required for the Lemma; recall Remark 1. In addition, the assumption holds if Σ_{m_0-1} is not positive semi-definite. Recall from Proposition 1 we have that Σ_{m_0-1} is not positive definite but, of course, it may still be positive semi-definite.

B.1 Theorem 1: Feasibility of the Estimation Method

Proof. When $K \geq m_0$ and $n > 2m_0$,

$$\{s_n \text{ exists}\} \supset \{s_n = m_0\} \supset \{s_n = m_0, t_n \geq K\}.$$

Then the result follows from Lemma 7. □

B.2 Theorem 2: Overestimation for the Asymptotic Generalized Variance of the Monte Carlo Error

Proof. We need to prove for all $\epsilon > 0$,

$$\Pr\left(\bigcap_{n=N}^{\infty} \{|\Sigma_{n,t_n}| > |\Sigma| - \epsilon\} \right) \to 1 \text{ as } N \to \infty. \quad (27)$$

Recall that Δ_i is defined as $\Delta_i = |\Sigma_i| - |\Sigma_{i-1}|$.

By Proposition 2(2) that $\lim_{m\to\infty} |\Sigma_m| = |\Sigma|$, we can write

$$\sum_{i=m_0+1}^{\infty} \Delta_i = |\Sigma| - |\Sigma_{m_0}| < \infty,$$

so $\sum_{i=m_0+1}^{\infty} \Delta_i$ converges; and hence the tail must converge to 0.

Therefore, for all $\epsilon > 0$, there exists $K_\epsilon \geq m_0$ such that

$$|\Sigma| - |\Sigma_{K_\epsilon}| = \sum_{i=K_\epsilon+1}^{\infty} \Delta_i < \epsilon/2. \tag{28}$$

Notice that

$$\begin{aligned} &\{|\Sigma_{n,t_n}| > |\Sigma| - \epsilon\} \\ \supset\ &\{|\Sigma_{n,t_n}| > |\Sigma| - \epsilon\} \bigcap \{s_n = m_0, t_n \geq K_\epsilon\} \\ \supset\ &\{|\Sigma_{n,K_\epsilon}| > |\Sigma| - \epsilon\} \bigcap \{s_n = m_0, t_n \geq K_\epsilon\}. \end{aligned} \tag{29}$$

The second step in (29) is due to the definition of mIS:

"$s_n = m_0$ and $t_n \geq K_\epsilon$ for some $K_\epsilon \geq m_0$" implies "$|\Sigma_{n,t_n}| \geq |\Sigma_{n,K_\epsilon}|$".

It follows directly from (29) that

$$\begin{aligned} &\bigcap_{n=N}^{\infty} \{|\Sigma_{n,t_n}| > |\Sigma| - \epsilon\} \\ \supset\ &\left(\bigcap_{n=N}^{\infty} \{|\Sigma_{n,K_\epsilon}| > |\Sigma| - \epsilon\}\right) \bigcap \left(\bigcap_{n=N}^{\infty} \{s_n = m_0, t_n \geq K_\epsilon\}\right). \end{aligned}$$

Therefore, to prove (27) it suffices to show

$$\Pr\left(\bigcap_{n=N}^{\infty} \{|\Sigma_{n,K_\epsilon}| > |\Sigma| - \epsilon\}\right) \to 1 \text{ as } N \to \infty, \tag{30}$$

and

$$\Pr\left(\bigcap_{n=N}^{\infty} \{s_n = m_0, t_n \geq K_\epsilon\}\right) \to 1 \text{ as } N \to \infty. \tag{31}$$

By the continuity of measure, (31) is equivalent to Lemma 7 and thus holds true. Then it remains to prove (30).

By Corollary 1 and 2, with probability 1,

$$|\Sigma_{n,K_\epsilon}| \to |\Sigma_{K_\epsilon}| \text{ as } n \to \infty,$$

which gives

$$\Pr\left(\bigcap_{n=N}^{\infty}\{\text{abs}(|\Sigma_{n,K_\epsilon}| - |\Sigma_{K_\epsilon}|) < \epsilon/2\}\right) \to 1 \text{ as } N \to \infty \tag{32}$$

where abs$(\cdot)$ denotes absolute value.

When $|\Sigma| - |\Sigma_{K_\epsilon}| < \epsilon/2$ as in (28),

$$\text{abs}(|\Sigma_{n,K_\epsilon}| - |\Sigma_{K_\epsilon}|) < \epsilon/2 \text{ implies } |\Sigma_{n,K_\epsilon}| > |\Sigma| - \epsilon,$$

so

$$\Pr\left(\bigcap_{n=N}^{\infty}\{\text{abs}(|\Sigma_{n,K_\epsilon}| - |\Sigma_{K_\epsilon}|) < \epsilon/2\}\right) \leq \Pr\left(\bigcap_{n=N}^{\infty}\{|\Sigma_{n,K_\epsilon}| > |\Sigma| - \epsilon\}\right). \tag{33}$$

Putting (32) and (33) together, we obtain (30). □

C Confidence Region with the Univariate Approach

We briefly state here the current methods for constructing confidence regions with univariate estimators. Let $\sigma(i)^2$ denote the (i,i)th entry of Σ. We treat the problem as p univariate cases, i.e., to estimate $\sigma(i)^2$ using univariate samples. Then we construct cube-shaped confidence regions.

Let $\mu_n(i)$ be the ith component of μ_n, and $\sigma_n(i)^2$ be the estimator for $\sigma(i)^2$.

The uncorrected confidence region is given by

$$C_n = \begin{Bmatrix} \mu_n(1) \pm z_{1-\alpha/2}\sigma_n(1)/\sqrt{n} \\ \mu_n(2) \pm z_{1-\alpha/2}\sigma_n(2)/\sqrt{n} \\ \vdots \\ \mu_n(p) \pm z_{1-\alpha/2}\sigma_n(p)/\sqrt{n} \end{Bmatrix}$$

with a volume of

$$\left(\frac{2z_{1-\alpha/2}}{\sqrt{n}}\right)^p \prod_{i=1}^{n}\sigma_n(i).$$

The Bonferroni confidence region for μ is

$$B_n = \begin{Bmatrix} \mu_n(1) \pm z_{1-\alpha/2p}\sigma_n(1)/\sqrt{n} \\ \mu_n(2) \pm z_{1-\alpha/2p}\sigma_n(2)/\sqrt{n} \\ \vdots \\ \mu_n(p) \pm z_{1-\alpha/2p}\sigma_n(p)/\sqrt{n} \end{Bmatrix}$$

with a volume of

$$\left(\frac{2z_{1-\alpha/2p}}{\sqrt{n}}\right)^p \prod_{i=1}^{n}\sigma_n(i).$$

References

[1] Q. D. Atkinson, R. D. Gray, and A. J. Drummond. mtDNA variation predicts population size in humans and reveals a major southern asian chapter in human prehistory. *Molecular Biology and Evolution*, 25:468–474, 2008. doi: 10.1093/molbev/msm277.

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

[3] C. R. Doss, J. M. Flegal, G. L. Jones, and R. C. Neath. Markov chain Monte Carlo estimation of quantiles. *Electronic Journal of Statistics*, 8:2448–2478, 2014.

[4] H. Doss and J. P. Hobert. Estimation of Bayes factors in a class of hierarchical random effects models using a geometrically ergodic MCMC algorithm. *Journal of Computational and Graphical Statistics*, 19:295–312, 2010. doi: 10.1198/jcgs.2010.09182.

[5] A. J. Drummond, S. Y. Ho, M. J. Phillips, and A. Rambaut. Relaxed phylogenetics and dating with confidence. *PLoS biology*, 4:e88, 2006. doi: 10.1371/journal.pbio.0040088.

[6] J. M. Flegal and G. L. Jones. Implementing MCMC: Estimating with confidence. In S. Brooks, A. Gelman, X.-L. Meng, and G. L. Jones, editors, *Handbook of Markov chain Monte Carlo*. Chapman & Hall, Boca Raton, 2011.

[7] J. M. Flegal, M. Haran, and G. L. Jones. Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260, 2008.

[8] B. E. Fristedt and L. F. Gray. *A Modern Approach to Probability Theory*. Birkhäuser, Boston, illustrated edition, 1996. ISBN 0817638075, 9780817638078.

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

[10] R. Giordano, T. Broderick, and M. Jordan. Linear response methods for accurate covariance estimates from mean field variational bayes. *arXiv:1506.04088*, 2015.

[11] O. Häggström and J. S. Rosenthal. On variance conditions for Markov chain CLTs. *Electronic Communications in Probability*, 12:454–464, 2007. doi: 10.1214/ECP.v12-1336.

[12] R. Harris, J. Beebe-Donk, H. Doss, and D. Burr. Aspirin, ibuprofen and other non-steroidal anti-inflammatory drugs in cancer prevention: a critical review of non-selective COX-2 blockade. *Oncology Reports*, 13:559–584, 2005.

[13] D. A. Harville. *Matrix Algebra from a Statistician's Perspective*. Springer Science & Business Media, New York, illustrated, reprint edition, 2008.

[14] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2, revised edition, 2012.

[15] A. A. Johnson and G. L. Jones. Geometric ergodicity of random scan Gibbs samplers for hierarchical one-way random effects models. *Journal of Multivariate Analysis*, 140: 325–342, 2015. doi: 10.1016/j.jmva.2015.06.002.

[16] A. A. Johnson, G. L. Jones, and R. C. Neath. Component-wise Markov chain Monte Carlo: Uniform and geometric ergodicity under mixing and composition. *Statistical Science*, 28:360–375, 2013.

[17] G. L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

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

[19] G. L. Jones, M. Haran, B. S. Caffo, and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547, 2006.

[20] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104:1–19, 1986. doi: 10.1007/BF01210789.

[21] M. R. Kosorok. Monte Carlo error estimation for multivariate Markov chains. *Statistics & Probability Letters*, 46:85–93, 2000. doi: 10.1016/S0167-7152(99)00090-5.

[22] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.

[23] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer London, illustrated edition, 1996.

[24] H. Ōsawa. Reversibility of first-order autoregressive processes. *Stochastic Processes and their Applications*, 28:61–69, 1988. doi: 10.1016/0304-4149(88)90064-6.

[25] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 1999.

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

[27] W. Rudin. *Functioanl Analysis*. McGraw-Hill, New York, second edition, 1991.

[28] D. Vats, J. M. Flegal, and G. L. Jones. Multivariate output analysis for Markov chain Monte Carlo. *Preprint arXiv:1512.07713*, 2015.

[29] D. Vats, J. M. Flegal, and G. L. Jones. Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Preprint arXiv:1507.08266*, 2016.