Journal of Multivariate Analysis 159 (2017) 184–199

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Multivariate initial sequence estimators in Markov chain Monte Carlo

Ning Dai *, Galin L. Jones

School of Statistics, University of Minnesota, United States

ARTICLE INFO

Article history:
Received 23 November 2016
Available online 9 June 2017

AMS 2010 subject classifications:
62H12

Keywords:
Markov chain Monte Carlo
Covariance matrix estimation
Central limit theorem
Metropolis–Hastings algorithm
Gibbs sampler

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 (1992) 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.

© 2017 Elsevier Inc. All rights reserved.

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 similar 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.

* Corresponding author.
E-mail addresses: daixx224@umn.edu (N. Dai), galin@umn.edu (G.L. Jones).

http://dx.doi.org/10.1016/j.jmva.2017.05.009
0047-259X/© 2017 Elsevier Inc. All rights reserved.

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,

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

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 $n \to \infty$,

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

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 [6,18,28].

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 (iid) 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

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

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 $P^n(x, B) = \Pr(X_{k+n} \in B \mid X_k = x)$, where Pr denotes probability. We assume that P satisfies detailed balance with respect to F. That is,

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

Download English Version:

https://daneshyari.com/en/article/5129390

Download Persian Version:

https://daneshyari.com/article/5129390

Daneshyari.com