Linear Algebra and its Applications 487 (2015) 184–202

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Constructing optimal transition matrix for Markov chain Monte Carlo

Sheng-Jhih Wu [a,*], Moody T. Chu [b,1]

[a] *Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou, China*
[b] *Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, United States*

ARTICLE INFO

Article history:
Received 30 May 2015
Accepted 8 September 2015
Available online 19 September 2015
Submitted by R. Brualdi

MSC:
15B51
49K35
60J10
65C05
65F18
90C26

Keywords:
MCMC
Asymptotic variance
Stationary distribution
Global optimization
Minimax problem
Inverse eigenvector problem
Stochastic matrix

ABSTRACT

The notion of asymptotic variance has been used as a means for gauging the performance of Markov chain Monte Carlo (MCMC) methods. For an effective MCMC simulation, it is imperative to first construct a Markov model with minimal asymptotic variance. The construction of such a stochastic matrix with prescribed stationary distribution as well as optimal asymptotic variance amounts to an interesting variationally constrained inverse eigenvector problem. Cast against a specially defined oblique coordinate system, the worst-case analysis of the asymptotic variance can be formulated as a problem of minimizing the logarithmic 2-norm of a restricted resolvent matrix over a convex and compact monoid. Based on this framework, this paper proposes employing global optimization techniques as a general instrument for numerical construction of optimal transition matrices. Numerical experiments manifest the complexity of the underlying problem. First, new global solutions different from the conventional structure characterized in the literature are found across the board for reversible problems. Second, local solutions with high frequencies of occurrence appear widespread for nonreversible problems. In all, the approach via the global

* Corresponding author.
E-mail addresses: szwu@suda.edu.cn (S.-J. Wu), chu@math.ncsu.edu (M.T. Chu).
[1] This research was supported in part by the National Science Foundation under grant DMS-1316779.

http://dx.doi.org/10.1016/j.laa.2015.09.016

optimization techniques is a feasible and practical means for numerical construction of the optimal Markov chain.

1. Introduction

For decades Markov chain Monte Carlo (MCMC) methods have been employed as a practical tool in a wide variety of applications such as Bayesian statistics, computational physics, genetics, and machine learning. See, for example, [3,13,24,26,27]. The methods become particularly useful when generating independent and identically distributed (i.i.d.) samples is not feasible or when the underlying distribution is not completely known. The basic idea underlying the MCMC is to construct a Markov chain with the desired distribution as its invariant distribution with the hope that, as the procedure runs long enough, the samples generated from the Markov chain serve as a good approximation to the would-be samples drawn from the unknown distribution. A key question to ask is how good an approximation is and the answer depends on the comparison criteria [2,4,19]. In the literature, one of the commonly employed measurements for gauging the performance of an MCMC algorithm is the so-called asymptotic variance which is the focus of this paper.

As a motivation, we briefly explain why the notion of asymptotic variance is a reasonable criterion for evaluating the performance of the MCMC methods. Let $S = \{1, 2, \cdots, n\}$ represent a finite state space and π be a probability distribution on S. It is often the case that we are interested in evaluating the expectation $\mathcal{E}(f) = \sum_{x \in S} f(x)\pi(x)$, where f is a real-valued function defined on S. When the closed-form calculation is not easy, we could appeal to the MCMC. Assume $X_0, X_1, \dots$ is a discrete time Markov chain on S with some transition probability matrix P and invariant distribution π. We then use the time average $\frac{1}{n}\sum_{i=0}^{n-1} f(X_i)$ as an estimation for the space average $\mathcal{E}(f)$, because by the strong law of large numbers we should have

$$\frac{\sum_{i=0}^{n-1} f(X_i)}{n} \xrightarrow{a.s.} \mathcal{E}(f). \tag{1}$$

In such a scenario, the asymptotic variance[2] is defined by

$$\nu(f, P) := \lim_{n\to\infty} n\mathcal{E}_{\mu_0}\left[\frac{\sum_{i=0}^{n-1} f(X_i)}{n} - \mathcal{E}(f)\right]^2, \tag{2}$$

[2] Strictly speaking, the expression in (2) measures the asymptotic mean squared error of the time average as an estimator for $\mathcal{E}(f)$. If the initial distribution μ_0 is precisely π, then the asymptotic mean squared error is equivalent to the asymptotic variance. In general, the initial distribution is biased, though might be of smaller order. The two notions are equivalent only if the limit distribution of $\lim_{n\to\infty} n\left(\frac{\sum_{i=0}^{n-1} f(X_i)}{n} - \mathcal{E}(f)\right)$ has zero mean, which is implicitly assumed in applications.