



On the asymptotic variance of reversible Markov chain without cycles

Chi-Hao Wu^a, Ting-Li Chen^{b,*}

^a Institute of Applied Mathematical Sciences, National Taiwan University, Taipei 10617, Taiwan

^b Institute of Statistical Sciences, Academia Sinica, Taipei 11529, Taiwan



ARTICLE INFO

Article history:

Received 18 October 2017

Received in revised form 16 January 2018

Accepted 19 January 2018

Available online 3 February 2018

Keywords:

Markov chain Monte Carlo

Rate of convergence

Reversibility

Asymptotic variance

ABSTRACT

Markov chain Monte Carlo (MCMC) is a popular approach to sample from high dimensional distributions, and the asymptotic variance is a commonly used criterion to evaluate the performance. While most popular MCMC algorithms are reversible, there is a growing literature on the development and analyses of nonreversible MCMC. Chen and Hwang (2013) showed that a reversible MCMC can be improved by adding an antisymmetric perturbation. They also raised a conjecture that it cannot be improved if there is no cycle in the corresponding graph. In this paper, we present a rigorous proof of this conjecture. The proof is based on the fact that the transition matrix with an acyclic structure will produce minimum commute time between vertices.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Markov chain Monte Carlo (MCMC) is a popular way to sample from high dimensional distributions with a wide variety of applications. For example, Diaconis (2009) provided several interesting examples including applications in cryptography and physics. Among the various applications of MCMC, one of the main purpose is to approximate the expectation of a specific function.

Let \mathcal{I} be a fixed index set of the finite state space, and π be a fixed probability measure on \mathcal{I} . The expectation of a real-valued function f with respect to π is denoted by $\pi(f) = \sum_{i \in \mathcal{I}} f(i)\pi(i)$. When the state space is extremely large, it is usually difficult to compute the expectation directly. In this case, we can construct a Markov chain with invariant distribution π . Then the empirical distribution of the samples $X_0, X_1, \dots, X_n, \dots$ converges to π . Therefore we can approximate $\pi(f)$ with the time average $\frac{1}{n} \sum_{k=0}^{n-1} f(X_k)$.

When we use MCMC algorithm with the transition matrix P to approximate the expectation of a function f , its asymptotic variance $v(f, P, \pi)$

$$v(f, P, \pi) = \lim_{n \rightarrow \infty} \mathbb{E}_\mu \left[\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right]^2,$$

where μ is an initial distribution, is a natural way to evaluate its performance. Note that the asymptotic variance is independent to the initial distribution (Iosifescu, 1980). Many researchers have studied how to optimize the performance of an MCMC algorithm based on this criterion. For examples, Peskun (1973) showed that the transition matrix with larger values in off-diagonal entries has a smaller asymptotic variance. Frigessi et al. (1992) derived the optimal reversible transition

* Corresponding author.

E-mail addresses: r05246009@ntu.edu.tw (C.-H. Wu), tlchen@stat.sinica.edu.tw (T.-L. Chen).

matrix under the worst-case scenario, while [Chen et al. \(2012\)](#) obtained the optimal transition matrix under the average case scenario. [Wu and Chu \(2015\)](#) proposed a global optimization technique in constructing the optimal transition matrix.

A necessary condition for the convergence of MCMC is that π is invariant under the transition P . However, the invariance is usually difficult to check when the state space is extremely large. The reversible condition, $\pi_i p_{i,j} = \pi_j p_{j,i}$, is a sufficient condition to guarantee the invariance, and most of the popular MCMC algorithms such as the Metropolis–Hastings algorithm ([Hastings, 1970](#); [Metropolis et al., 1953](#)) and the Gibbs sampler ([Geman and Geman, 1984](#)) make use of reversibility. Though the reversible condition makes the construction of a transition matrix easier, researchers have shown that reversible chains may not be as efficient as the non-reversible ones ([Diaconis et al., 2000](#); [Mira and Geyer, 2000](#); [Hwang et al., 1993, 2005](#)). Recently, there is an increasing number of works on non-reversible MCMC ([Chen et al., 2012](#); [Chen and Hwang, 2013](#); [Bierkens, 2016](#); [Bierkens and Roberts, 2017](#); [Poncet, 2017](#)).

Since reversible transition matrices are found to be less efficient, two interesting questions arise naturally. What kinds of reversible transition matrix can be improved? How can we improve these reversible transition matrices? An earlier result can be found in [Peskun \(1973\)](#), and [Chen and Hwang \(2013\)](#) showed that the chain can be uniformly better by adding an antisymmetric perturbation when the corresponding graph has cycles. They also proposed a conjecture that if the graph corresponding to the reversible transition matrix does not have any cycle, there does not exist a Markov chain that is uniformly better.

In this paper, we give a proof to the conjecture. It starts with some preliminaries listed in Section 2, and the main result is presented in Section 3. Finally, the conclusion is summarized in Section 4.

2. Preliminaries

In this paper, we only consider irreducible Markov chain. We start with introducing some notations and terminologies that will be used throughout the text.

A reversible Markov chain with a transition matrix P has a corresponding undirected graph $G = (V, E)$, where V is its state space and E is the set of its edges. For any pair of states i and j , $p_{i,j} > 0$ if and only if $(i, j) \in E$.

A cycle in a graph $G = (V, E)$ is a path which starts and ends at the same vertex without any other repeated vertices along the path. When the graph $G = (V, E)$ is undirected, a cycle must contain at least 3 different vertices. If a graph has no cycle, we say it is acyclic.

An undirected acyclic graph $G = (V, E)$ has a tree structure. We say a vertex i is a leaf if there is only one edge (i, j) in E that directly connects i with another vertex j .

Definition 1. For real-valued functions f and g on \mathcal{I} , their inner product with respect to a probability distribution π is defined as

$$\langle f, g \rangle_\pi = \sum_{i \in \mathcal{I}} f(i)g(i)\pi(i).$$

Lemma 1. For any real-valued function f

$$v(f, P, \pi) = 2\langle [(I - P + \Pi)^{-1} - \Pi] \{f - \pi(f)\}, f - \pi(f) \rangle_\pi - \langle f - \pi(f), f - \pi(f) \rangle_\pi$$

where Π is a matrix whose rows all equal π .

Proof. See Theorem 4.8 in [Iosifescu \(1980\)](#). \square

When we consider only real-valued function in $\mathcal{N} = \{f : \pi(f) = 0\}$, the formula will be simplified into $v(f, P, \pi) = 2\langle [(I - P + \Pi)^{-1} - \Pi]f, f \rangle_\pi - \langle f, f \rangle_\pi$. $(I - P + \Pi)^{-1} - \Pi$ is the so called fundamental matrix, and we denote it as Z . Since $\langle f, f \rangle_\pi$ is independent to the transition matrix, the minimization of the asymptotic variance is equivalent to the minimization of $\langle Zf, f \rangle_\pi$.

The first hitting time and the first return time are defined as

$$T_i = \inf\{t \geq 0 : X_t = i\}$$

and

$$T_i^+ = \inf\{t \geq 1 : X_t = i\}.$$

In the following, we list some useful formulas that will be used in our proof.

Lemma 2.

$$z_{ij} = \pi_j \mathbb{E}_\pi T_j - \pi_i \mathbb{E}_i T_j.$$

Proof. See Lemma 11 and Lemma 12 in Chapter 2 of [Aldous and Fill \(2002\)](#). \square

Download English Version:

<https://daneshyari.com/en/article/7548378>

Download Persian Version:

<https://daneshyari.com/article/7548378>

[Daneshyari.com](https://daneshyari.com)