Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 123 (2013) 3861–3876

stochastic processes and their applications

www.elsevier.com/locate/spa

Analysis of convergence rates of some Gibbs samplers on continuous state spaces

Aaron Smith[*,1]

Department of Mathematics, Stanford University, Stanford, CA 94305, United States

Received 29 June 2012; received in revised form 15 February 2013; accepted 11 May 2013
Available online 18 May 2013

Abstract

We use a non-Markovian coupling and small modifications of techniques from the theory of finite Markov chains to analyze some Markov chains on continuous state spaces. The first is a generalization of a sampler introduced by Randall and Winkler, and the second a Gibbs sampler on narrow contingency tables.

Keywords: Gibbs sampler; Markov chain; Mixing time; Coupling

1. Introduction

The problem of sampling from a given distribution on high-dimensional continuous spaces arises in the computational sciences and Bayesian statistics, and a frequently-used solution is Markov chain Monte Carlo (MCMC); see [14] for many examples. Because MCMC methods produce good samples only after a lengthy mixing period, a long-standing mathematical question is to analyze the mixing times of the MCMC algorithms which are in common use. Although there are many mixing conditions, the most commonly used is called the mixing time, and is based on the total variation distance.

For measures ν, μ with common measurable σ-algebra $\mathcal{A}$, the *total variation distance* between μ and ν is

$$\|\mu - \nu\|_{TV} = \sup_{A \in \mathcal{A}} (\mu(A) - \nu(A)).$$

* Correspondence to: 2440 Wyndale Cr, K1H 7A6 Ottawa, ON, Canada. Tel.: +1 6509068914.
E-mail addresses: asmith3@math.stanford.edu, smith.aaron.matthew@gmail.com.
[1] Present address: ICERM, Brown University, Providence RI, United States.

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.05.003

For an ergodic discrete-time Markov chain X_t with unique stationary distribution π, the *mixing time* is

$$\tau(\epsilon) = \inf\{t : \|\mathcal{L}(X_t) - \pi\|_{TV} < \epsilon\}.$$

Although most scientific and statistical uses of MCMC methods occur in continuous state spaces, much of the mathematical mixing analysis has been in the discrete setting. The methods that have been developed for discrete chains often break down when used to analyze continuous chains, though there are some efforts, such as [26,22,16], to create general techniques. This paper extends the author's previous work in [25] and work of Randall and Winkler [20], and attempts to provide some more examples of relatively sharp analyses of continuous chains similar to those used to develop the discrete theory.

The first process that we analyze is a Gibbs sampler on the simplex with a very restricted set of allowed moves. Fix a finite group G of size n with symmetric generating set R of size m, with $id \notin R$. For unity of notation, label the group elements with the integers from 1 to n. We consider the process $X_t[g]$ on the simplex $\Delta_G = \{X \in \mathbb{R}^n \mid \sum_{g \in G} X[g] = 1; X[g] \geq 0\}$. At each step, choose $g \in G$, $r \in R$ and $\lambda \in [0, 1]$ uniformly, and set

$$\begin{aligned} X_{t+1}[g] &= \lambda(X_t[g] + X_t[gr]) \\ X_{t+1}[gr] &= (1 - \lambda)(X_t[g] + X_t[gr]). \end{aligned} \tag{1}$$

For all other $h \in G$ set $X_{t+1}[h] = X_t[h]$. Let U_G be the uniform distribution on Δ_G; this is also the stationary distribution of X_t. Also consider a random walk Z_t on G, where in each stage we choose $g \in G$ and $r \in R$ uniformly at random and set $Z_{t+1} = gr$ if $Z_t = g$, set $Z_{t+1} = g$ if $Z_t = gr$, and $Z_{t+1} = Z_t$ otherwise. This is the standard simple random walk on the Cayley graph, slowed down by a factor of about n. Let $\widehat{\gamma}$ be the spectral gap of the walk Z_t, and follow the notation that $\mathcal{L}(X)$ denotes the distribution of a random variable X.

Theorem 1 (*Convergence Rate for Gibbs Sampler with Geometry*)**.** *For $T > \frac{8C}{\widehat{\gamma}} \log(n)$, $C > \frac{103}{4}$, and n satisfying $n > \max\left(4096, \frac{C}{3} + \frac{10}{3}, \left(\frac{C}{3} - \frac{13}{12}\right) \log(n)\right)$*

$$\|\mathcal{L}(X_T) - U_G\|_{TV} \leq 7n^{4.5 - \frac{C}{6}}$$

and conversely for $T < \frac{k}{\widehat{\gamma}}$,

$$\|\mathcal{L}(X_T) - U_G\|_{TV} \geq \frac{1}{2} e^{-k} - 4n^{-\frac{1}{3}}.$$

This substantially generalizes [20,25], from samplers corresponding to $G = \mathbb{Z}_n$, and $R = \{1, -1\}$ or $R = \mathbb{Z}_n \setminus \{0\}$ respectively, to general Cayley graphs. In addition to being of mathematical interest, this process is an example of a gossip process with some geometry, studied by electrical engineers and sociologists interested in how information propagates through networks; see [23] for a survey.

The proof of the upper bound will use an auxiliary chain similar to that found in [20], a coupling argument improved from [25], and an unusual use of comparison theory from [7]. The proof of the lower bound is elementary.

The next example consists of narrow contingency tables. Beginning with the work of Diaconis and Efron [4] on independence tests, there has been interest in finding efficient ways to sample uniformly from the collection of integer-valued matrices with given row and column sums.