Statistics and Probability Letters 91 (2014) 110–116

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On reparametrization and the Gibbs sampler

Jorge Carlos Román [a,*], James P. Hobert [b], Brett Presnell [b]

[a] Department of Mathematics, Vanderbilt University, United States

[b] Department of Statistics, University of Florida, United States

ARTICLE INFO

Article history:
Received 17 January 2014
Received in revised form 19 March 2014
Accepted 23 March 2014
Available online 12 April 2014

Keywords:
Markov chain
Geometric ergodicity
Monte Carlo
Convergence rate
Parametrization

ABSTRACT

Gibbs samplers derived under different parametrizations of the target density can have radically different rates of convergence. In this article, we specify conditions under which reparametrization leaves the convergence rate of a Gibbs chain unchanged. An example illustrates how these results can be exploited in convergence rate analyses.

1. Introduction

It is well-known that Gibbs samplers derived under different parametrizations of a Bayesian hierarchical model can have dramatically different rates of convergence (Gelfand et al., 1995; Roberts and Sahu, 1997; Papaspiliopoulos et al., 2007; Yu and Meng, 2011). In this article, we consider the reverse situation in which reparametrization has no effect. To motivate our study, we begin with a fresh look at a well-known toy example involving a simple random effects model with known variance components.

Consider the one-way random effects model given by

$$Y_{ij} = \theta_i + \epsilon_{ij}, \tag{1}$$

$i = 1, \dots, c, j = 1, \dots, m_i$, where the θ_is are independent and identically distributed (iid) $\mathrm{N}(\mu, \sigma^2)$, and the ϵ_{ij}s are independent of the θ_is and iid $\mathrm{N}(0, \sigma_e^2)$. (For now, we restrict attention to the balanced case where $m_i \equiv m$.) Suppose that the variance components, σ^2 and σ_e^2, are known and that the prior on μ is flat. Let $\theta = (\theta_1, \dots, \theta_c)$ and let y denote the observed data. A simple calculation shows that the posterior density of μ given y is normal, but consider nevertheless the two-component Gibbs chain $\{(\mu_n, \theta_n)\}_{n=0}^{\infty}$ that alternately samples from the conditional distributions $\theta|\mu, y$ and $\mu|\theta, y$, which are c-variate normal and univariate normal, respectively. The marginal sequence $\{\mu_n\}_{n=0}^{\infty}$ is itself a Markov chain whose invariant density is the posterior density (of μ given y), and it is easy to show that the exact rate of convergence of this chain is $\sigma_e^2/(\sigma_e^2 + m\sigma^2)$ (see, e.g., Liu et al., 1994). The "rate of convergence" will be formally defined in Section 2, but for now it suffices to note that the rate is between 0 and 1, and smaller is better.

Now consider a reparametrized version of model (1) given by $Y_{ij} = \mu + u_i + \epsilon_{ij}$, where the u_is are iid $\mathrm{N}(0, \sigma^2)$, and the ϵ_{ij}s are independent of the u_is and still iid $\mathrm{N}(0, \sigma_e^2)$. Let $u = (u_1, \dots, u_c)$. This is called the *non-centered parametrization* (NCP),

* Corresponding author. Tel.: +1 615 322 3754.
E-mail address: jc.roman@vanderbilt.edu (J.C. Román).

http://dx.doi.org/10.1016/j.spl.2014.03.024

whereas model (1) is called the *centered parametrization* (CP). If we put the same flat prior on μ, then the posterior density of μ given y remains the same as in the CP model. However, the two-component Gibbs sampler derived from the NCP model, which alternates between draws from $u|\mu, y$ and $\mu|u, y$, is not the same as the one based on the CP. Furthermore, the two Gibbs samplers have completely different convergence behaviors. Indeed, the convergence rate of the NCP Gibbs sampler is $1-\sigma_e^2/(\sigma_e^2+m\sigma^2)$. So when one of the two Gibbs samplers is very slow to converge, the other converges extremely rapidly. This simple example illustrates that reparametrization can significantly affect the convergence rate of the Gibbs sampler.

In a practical version of the one-way model, the variance components are unknown. In this case, the standard default prior density for $(\mu, \sigma^2, \sigma_e^2)$ is $1/(\sigma_e^2\sqrt{\sigma^2})$. We assume that the posterior is proper — see Román (2012) for conditions. The posterior density of $(\mu, \sigma^2, \sigma_e^2)$ given y, which is the same under CP and NCP, is intractable, so this is no longer a toy example. As in the known variance case, there are two different versions of the standard two-component Gibbs sampler for this problem: the CP Gibbs sampler, which alternates between $\theta, \mu|\sigma^2, \sigma_e^2, y$ and $\sigma^2, \sigma_e^2|\mu, \theta, y$, and the NCP Gibbs sampler, which alternates between $u, \mu|\sigma^2, \sigma_e^2, y$ and $\sigma^2, \sigma_e^2|u, \mu, y$. The results of Section 3 imply that, in contrast with the known variance case, these two Gibbs samplers converge at exactly the same rate. Consequently, convergence rate results for either of these Gibbs samplers apply directly to the other. In Section 3 we compare the results of Román (2012), who analyzed the NCP Gibbs sampler, with those of Tan and Hobert (2009), who studied the CP version.

The CP and NCP Gibbs Markov chains described above share the same rate of convergence because the transformation that takes the CP model to the NCP model involves variables (θ and μ) that reside in the same component (or block) of the two-component Gibbs sampler. (Note that this is not the case in the toy example where the variance components are known.) The main result in this paper is a formalization of this idea. We now provide an overview of our results in the special case where the target distribution has a density with respect to Lebesgue measure.

Suppose $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_k} \to [0, \infty)$ is a probability density function, and let $\Phi_1 = \left\{(X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(k)})\right\}_{n=0}^{\infty}$ denote the Markov chain simulated by the k-component Gibbs sampler based on $f(x_1, x_2, \ldots, x_k)$ that updates the components in the natural order. It is well-known and easy to see that the marginal sequence $\Phi_{-1} := \left\{(X_n^{(2)}, \ldots, X_n^{(k)})\right\}_{n=0}^{\infty}$ is also a Markov chain. Now, for $i \in \{2, 3, \ldots, k\}$, let Φ_i denote the k-component Gibbs sampler whose update order is $(i, i+1, \ldots, k, 1, 2, \ldots, i-1)$, and let Φ_{-i} denote the corresponding marginal Markov chain (that leaves out $X^{(i)}$). We show that all $2k$ of these chains converge at exactly the same rate. Not only is this fact the key to the proof of our main result concerning reparametrization, it is also useful from a practical standpoint. Indeed, if one wishes to know the rate of convergence of Φ_1, then it suffices to study the lower-dimensional chain Φ_{-i} (for any $i = 1, 2, \ldots, k$), which may be easier to analyze than Φ_1. This idea has been used to establish qualitative convergence results (such as geometric and uniform ergodicity) for two-component Gibbs samplers (see, e.g., Diebolt and Robert (1994) and Román and Hobert (2012)).

Now let $(X_1, X_2, \ldots, X_k)$ denote a random vector with density f, and consider the k-component Gibbs sampler $\tilde{\Phi}_1$ based on the distribution of $(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_k) = (t_1(X_1), t_2(X_2), \ldots, t_k(X_k))$. Suppose $f(x_1, x_2, \ldots, x_k)$ can be written as a function of $(t_1(x_1), t_2(x_2), \ldots, t_k(x_k))$, an assumption that obviously holds if each $t_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i}$ is invertible. Then, by exploiting the fact that the $2k$ chains described above share the same rate, we show that Φ_1 and $\tilde{\Phi}_1$ converge at the same rate. An important implication of this result is that, when analyzing the convergence rate of a Gibbs sampler, one is free to choose a convenient parametrization, as long as the corresponding transformation respects the "within-component" restriction.

The remainder of this article is organized as follows. Section 2 contains some background on general state space Markov chain theory as well as preliminary results. Our main result showing that a within-component reparametrization does not affect the convergence rate of the Gibbs Markov chain can be found in Section 3. This section also contains the application of our main result to the Gibbs samplers for the one-way model with improper priors.

2. Markov chain background and preliminary results

As in Meyn and Tweedie (1993, Chapter 3), let $P(x, dy)$ be a generic Markov transition function (*MTF*) on a set X equipped with a countably generated σ-algebra. Let $P^n(x, dy)$ denote the n-step MTF. We assume throughout that the chain determined by P is ψ-irreducible, aperiodic and positive recurrent with invariant probability measure π. We *do not* assume reversibility.

For a measure ν on X, let $\nu P^n(dy) = \int_{\mathsf{X}} P^n(x, dy)\nu(dx)$. Following Roberts and Tweedie (2001) and Rosenthal (2003), define the L^1-rate of convergence of the Markov chain as

$$\rho = \exp\left\{\sup_{\nu\in p(\pi)} \lim_{n\to\infty} \frac{1}{n} \log \|\nu P^n - \pi\|_{\text{TV}}\right\},$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm for signed measures and $p(\pi)$ is the set of all probability measures ν that are absolutely continuous with respect to π with $\int_{\mathsf{X}} (d\nu/d\pi)^2\, d\pi < \infty$. For reversible chains, ρ equals the "usual" rate of convergence, i.e., the spectral radius (and norm) of the self-adjoint Markov operator defined by P (Rosenthal, 2003, Proposition 2).

As in Roberts and Rosenthal (1997), we say that the chain (or the corresponding MTF) is π-a.e. geometrically ergodic if there exist $M : \mathsf{X} \to (0, \infty)$ and $\kappa < 1$ such that, for π-a.e. $x \in \mathsf{X}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{\text{TV}} \le M(x)\kappa^n \quad \text{for all } n \in \mathbb{N}.$$