Computational Statistics and Data Analysis 78 (2014) 141–152

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A modified conditional Metropolis–Hastings sampler

Alicia A. Johnson [a,*], James M. Flegal [b]

[a] Department of Mathematics, Statistics, and Computer Science, Macalester College, United States

[b] Department of Statistics, University of California, Riverside, United States

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form 10 April 2014
Accepted 14 April 2014
Available online 24 April 2014

Keywords:
Markov chain Monte Carlo
Metropolis–Hastings
Gibbs sampler
Geometric ergodicity

ABSTRACT

A modified conditional Metropolis–Hastings sampler for general state spaces is introduced. Under specified conditions, this modification can lead to substantial gains in statistical efficiency while maintaining the overall quality of convergence. Results are illustrated in two settings: a toy bivariate Normal model and a Bayesian version of the random effects model.

1. Introduction

Consider a random variable $X = (X_1, \ldots, X_m)$ where $X_i \in \mathbb{R}^{d_i}$ for $i = 1, \ldots, m$ and $d_i \geq 1$. Let X have probability distribution ϖ with support $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and associated conditional distributions $\varpi_{X_i|X_{-i}}$ where $X_{-i} = X \setminus X_i$. Further, with respect to measure $\mu = \mu_1 \times \cdots \times \mu_m$, suppose that ϖ admits density $\pi(x_1, \ldots, x_m)$ with associated full conditional densities $\pi(x_i|x_{-i})$. When ϖ is intractable, inference regarding X may require Markov chain Monte Carlo (MCMC) methods. To this end, consider using the conditional Metropolis–Hastings algorithm (CMH) under a random scan to construct a Markov chain denoted as

$$\Phi = \left\{X^{(0)}, X^{(1)}, \ldots\right\} = \left\{\left(X_1^{(0)}, \ldots, X_m^{(0)}\right), \left(X_1^{(1)}, \ldots, X_m^{(1)}\right), \ldots\right\}.$$

Under a fixed set of probabilities $p = (p_1, \ldots, p_m)$ where $0 < p_i < 1$ and $\sum_{i=1}^m p_i = 1$, Φ moves from $X^{(i)} = x$ to $X^{(i+1)}$ by updating a single randomly selected X_i while fixing all others. Specifically, in iteration $j+1$ of the CMH, first draw $(Z_1, \ldots, Z_m) \sim \text{Multinomial}(1, p)$. Then for $\{i : Z_i = 1\}$, draw x_i' from a proposal density $\tilde{q}_i(x_i'|x)$ and replace x_i with x_i' with acceptance probability

$$\tilde{\alpha}_i(x_i'|x) = \min\left\{1, \frac{\pi_i(x_i'|x_{-i})}{\pi_i(x_i|x_{-i})} \frac{\tilde{q}_i(x_i|x_{[i-1]}, x_i', x^{[i+1]})}{\tilde{q}_i(x_i'|x)}\right\}$$

where $x_{[i]} = (x_1, \ldots, x_i)$ and $x^{[i]} = (x_i, \ldots, x_m)$. The Gibbs sampler (GS) is a special case of the CMH with $\tilde{q}_i(x_i|x) = \pi_i(x_i|x_{-i})$ for all i when the latter are tractable.

After n iterations, we can estimate the expected value $\beta := E_\varpi f = \int f(x)\varpi(dx)$ of some function of interest, $f : \mathcal{X} \to \mathbb{R}$, by the Monte Carlo average $\hat{\beta}_n := \frac{1}{n}\sum_{i=0}^{n-1} f\left(X^{(i)}\right)$. The level of confidence we can place in $\hat{\beta}_n$ is intimately tied to the rate at

* Corresponding author.
E-mail addresses: ajohns24@macalester.edu (A.A. Johnson), jflegal@ucr.edu (J.M. Flegal).

http://dx.doi.org/10.1016/j.csda.2014.04.009
0167-9473/

which Φ converges to ϖ. To this end, assume that Φ is Harris ergodic (Meyn and Tweedie, 1993) and define n-step transition kernel $P^n(x, A) = \Pr\left(X^{(i+n)} \in A \; X^{(i)} = x\right)$ for $x \in \mathcal{X}$, $n, i \in \mathbb{N}$, and $A \in \mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra associated with $\mathcal{X}$. Then we say Φ is *geometrically ergodic* if it converges to ϖ in total variation distance at a geometric rate. That is, there exist function $M : \mathcal{X} \to \mathbb{R}$ and constant $t \in (0, 1)$ such that

$$\|P^n(x, \cdot) - \varpi(\cdot)\|_{\text{TVD}} := \sup_{A\in\mathcal{B}} |P^n(x, A) - \varpi(A)| \le t^n M(x) \quad \text{for all } x \in \mathcal{X}.$$

In addition to guaranteeing effective simulation results in finite time, geometric ergodicity is a key sufficient condition for the existence of a Markov chain central limit theorem for $\hat{\beta}_n$ (Jones, 2004).

Inspired by the work of Liu (1996) for the GS on discrete state spaces, we show that a simple modification to the CMH (hence GS) can lead to significant improvements in the Markov chain efficiency and quality of estimates $\hat{\beta}_n$. Specifically, we introduce a modified CMH (MCMH) that increases efficiency by encouraging movement of $X_j^{(i+1)}$ outside the local neighborhood of $X_j^{(i)}$, denoted by $B_j \subset \mathcal{X}_j$. We show that this modification maintains the overall quality of convergence; geometric ergodicity of the MCMH guarantees the same for the CMH and, under conditions on B_j, the reverse is also true.

Further, we explore the impact of B_j on the MCMH and compare the empirical performance of the CMH and MCMH in two different model settings: (1) a bivariate Normal model, and (2) a Bayesian version of the random effects model. The latter is practically relevant in that inference for this model *requires* MCMC methods. In both settings, the MCMH with reasonably sized B_j is significantly more efficient than the CMH in both its movement around state space $\mathcal{X}$ and in its estimation of expected value β. However, there are limits to the MCMH efficiency. Mainly, when B_j are too large, the MCMH requires significantly more computational effort and is pushed out to the 'edges' of the state space. Thus, in these settings, the MCMH cannot compete with the CMH.

Our paper is organized as follows. Section 2 introduces the MCMH and compares convergence among the CMH and MCMH. Section 3 explores these results with applications in two model settings. All proofs are deferred to the Appendix.

2. The modified CMH algorithm

Consider a Markov chain for target density $\pi(x_1, \ldots, x_m)$ with associated full conditionals $\pi(x_i|x_{-i})$. The transition kernel of the CMH for π with the proposal densities $\tilde{q}_i(x_i'|x)$ and acceptance probabilities $\tilde{\alpha}_i(x_i'|x)$ outlined above can be expressed as

$$P_{CMH}(x, A) = \sum_{i=1}^{m} p_i P_{CMH_i}(x, A)$$

where P_{CMH_i}, Markov kernels corresponding to the X_i updates, are defined by

$$P_{CMH_i}(x, A) = \int_{\{x_i' : (x_{[i-1]}, x_i', x^{[i+1]}) \in A\}} \tilde{q}_i(x_i'|x)\tilde{\alpha}_i(x_i'|x)\mu_i(dx_i') + \left[1 - \int \tilde{q}_i(x_i'|x)\tilde{\alpha}_i(x_i'|x)\mu_i(dx_i')\right] I(x \in A).$$

Ideally, the CMH will tour all reaches of $\mathcal{X}$ without getting stuck for too long in any one "corner". Indeed, we can facilitate such movement with a simple modification to the CMH algorithm. Letting x denote the current state of the CMH, suppose that component x_i is selected for update and let $B_i(x_i|x_{-i}) \subset \mathcal{X}_i$ be a local neighborhood of x_i that could depend on x_{-i}. For example, we might define $B_i(x_i|x_{-i}) = x_i \pm \varepsilon$ for $\varepsilon > 0$ when $\mathcal{X}_i = \mathbb{R}$ or define $B_i(x_i|x_{-i})$ to be a circle centered at x_i with radius ε when $\mathcal{X}_i = \mathbb{R}^2$. Then instead of proposing an x_i update from $\tilde{q}(\cdot|x)$, we can restrict movement to states outside $B_i(x_i|x_{-i})$, i.e. $B_i^c(x_i|x_{-i}) = \mathcal{X}_i \setminus B_i(x_i|x_{-i})$, through a Metropolis–Hastings step as follows. First draw x_i' from the proposal density

$$q_i(x_i'|x) = \frac{\tilde{q}_i(x_i'|x)}{\int_{B_i^c(x_i|x_{-i})} \tilde{q}_i(z_i|x)\mu_i(dz_i)} I(x_i' \in B_i^c(x_i|x_{-i})),$$

for example, using an accept–reject strategy, and then replacing x_i with x_i' with acceptance probability

$$\alpha_i(x_i'|x) = \min\left\{1,\ \frac{\pi_i(x_i'|x_{-i})}{\pi_i(x_i|x_{-i})} \frac{\tilde{q}_i(x_i|x_{[i-1]}, x_i', x^{[i+1]}) \int_{B_i^c(x_i|x_{-i})} \tilde{q}_i(z_i|x)\mu_i(dz_i)}{\tilde{q}_i(x_i'|x) \int_{B_i^c(x_i'|x_{-i})} \tilde{q}_i(z_i|x)\mu_i(dz_i)}\right\}.$$

Thus, the modified CMH (MCMH) has transition kernel

$$P_{MCMH}(x, A) = \sum_{i=1}^{m} p_i P_{MCMH_i}(x, A)$$

for

$$P_{MCMH_i}(x, A) = \int_{\{x_i' : (x_{[i-1]}, x_i', x^{[i+1]}) \in A\}} q_i(x_i'|x)\alpha_i(x_i'|x)\mu_i(dx_i') + \left[1 - \int q_i(x_i'|x)\alpha_i(x_i'|x)\mu_i(dx_i')\right] I(x \in A).$$

Note the dependence of the MCMH on neighborhoods B_i. If $B_i(x_i|x_{-i}) = \emptyset$ for all i, the MCMH and CMH are equivalent. At the other extreme, when $B_i(x_i|x_{-i}) = \mathcal{X}_i$, the MCMH Markov chain has nowhere to move. Thus we restrict our attention to