



Trace-class Monte Carlo Markov chains for Bayesian multivariate linear regression with non-Gaussian errors

Qian Qin ^{*}, James P. Hobert

Department of Statistics, University of Florida, United States



ARTICLE INFO

Article history:

Received 12 January 2017

Available online 5 April 2018

AMS subject classifications:

60J05

Keywords:

Compact operator

Data augmentation algorithm

Haar PX-DA algorithm

Heavy-tailed distribution

Scale mixture

Markov operator

Trace-class operator

ABSTRACT

Let π denote the intractable posterior density that results when the likelihood from a multivariate linear regression model with errors from a scale mixture of normals is combined with the standard non-informative prior. There is a simple data augmentation algorithm (based on latent data from the mixing density) that can be used to explore π . Let h and d denote the mixing density and the dimension of the regression model, respectively. Hobert et al. (2018) have recently shown that, if h converges to 0 at the origin at an appropriate rate, and $\int_0^\infty u^{d/2} h(u) du < \infty$, then the Markov chains underlying the data augmentation (DA) algorithm and an alternative Haar parameter expanded DA (PX-DA) algorithm are both geometrically ergodic. Their results are established using probabilistic techniques based on drift and minorization conditions. In this paper, spectral analytic techniques are used to establish that something much stronger than geometric ergodicity often holds. In particular, it is shown that, under simple conditions on h , the Markov operators defined by the DA and Haar PX-DA Markov chains are *trace-class*, i.e., compact with summable eigenvalues. Many standard mixing densities satisfy the conditions developed in this paper. Indeed, the new results imply that the DA and Haar PX-DA Markov operators are trace-class whenever the mixing density is generalized inverse Gaussian, log-normal, Fréchet (with shape parameter larger than $d/2$), or inverted Gamma (with shape parameter larger than $d/2$).

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Consider the multivariate linear regression model

$$Y = X\beta + \varepsilon \Sigma^{1/2}, \quad (1)$$

where Y denotes an $n \times d$ matrix of responses, X is an $n \times p$ matrix of known covariates, β is a $p \times d$ matrix of unknown regression coefficients, $\Sigma^{1/2}$ is an unknown positive-definite scale matrix, and ε is an $n \times d$ matrix whose rows are iid random vectors from a scale mixture of multivariate normal densities. In particular, letting ε_i^\top denote the i th row of ε , we assume that ε_i has density

$$f_h(\varepsilon_i) = \int_0^\infty \frac{u^{d/2}}{(2\pi)^{d/2}} \exp\left(-\frac{u}{2} \varepsilon_i^\top \varepsilon_i\right) h(u) du,$$

^{*} Corresponding author.

E-mail address: qianqin11@ufl.edu (Q. Qin).

where $h : (0, \infty) \rightarrow [0, \infty)$ is the so-called *mixing density*. Error densities of this form are often used when heavy-tailed errors are required. For example, it is well known that if h is a $\mathcal{G}(\nu/2, \nu/2)$ gamma density (with mean 1), then f_h becomes the multivariate Student's t density with ν degrees of freedom.

A Bayesian analysis of the data from this regression model requires a prior on (β, Σ) . We consider an improper default prior that takes the form $\omega(\beta, \Sigma) \propto |\Sigma|^{-a} \mathbf{1}_{S_d}(\Sigma)$, where $S_d \subset \mathbb{R}^{d(d+1)/2}$ denotes the space of $d \times d$ positive definite matrices. Taking $a = (d + 1)/2$ yields the independence Jeffreys prior, which is the standard non-informative prior for multivariate location scale problems. Of course, whenever an improper prior is used, one must check that the corresponding posterior distribution is proper. Letting y denote the observed value of Y , the joint density of the data from Model (1) can be expressed as

$$f(y|\beta, \Sigma) = \prod_{i=1}^n \left[\int_0^\infty \frac{u^{d/2}}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{u}{2} (y_i - \beta^\top x_i)^\top \Sigma^{-1} (y_i - \beta^\top x_i) \right\} h(u) du \right].$$

Define

$$m(y) = \int_{S_d} \int_{\mathbb{R}^{p \times d}} f(y|\beta, \Sigma) \omega(\beta, \Sigma) d\beta d\Sigma.$$

The posterior distribution is proper precisely when $m(y) < \infty$. Let Λ stand for the $n \times (p + d)$ matrix $(X : y)$. Straightforward arguments using ideas from Fernández and Steel [6] show that, together, the following four conditions are sufficient for posterior propriety:

- (S1) $\text{rank}(\Lambda) = p + d$;
- (S2) $n > p + 2d - 2a$;
- (S3) $\int_0^\infty u^{d/2} h(u) du < \infty$;
- (S4) $\int_0^\infty u^{-(n-p+2a-2d-1)/2} h(u) du < \infty$.

These four conditions are assumed to hold throughout this paper.

Remark 1. Conditions (S1) and (S2) are known to be necessary for posterior propriety [6,9].

Remark 2. Condition (S3) clearly concerns the tail behavior of h . Similarly, condition (S4) concerns the behavior of h near the origin, unless $n - p + 2a - 2d - 1$ is negative, which is possible. Note, however, that (S2) implies that $-(n - p + 2a - 2d - 1)/2 < 1/2$. Consequently, if $n - p + 2a - 2d - 1$ is negative, then (S4) is implied by (S3).

Of course, the posterior density of (β, Σ) given the data takes the form

$$\pi(\beta, \Sigma|y) = \frac{f(y|\beta, \Sigma) \omega(\beta, \Sigma)}{m(y)}.$$

There is a well-known data augmentation (DA) algorithm that can be used to explore this intractable density [15]. Hobert et al. [9] (hereafter HJKQ) performed convergence rate analyses of the Markov chains underlying this DA algorithm and an alternative Haar parameter expanded DA (PX-DA) algorithm.

In this paper, we examine the chains from an operator theory perspective, and provide a substantial improvement of HJKQ's main result. A formal statement of the DA algorithm requires some buildup; unless otherwise noted, "the DA algorithm" refers to the algorithm of Liu [15], rather than its Haar PX-DA alternative, which will be defined in Section 4.

Let $z = (z_1, \dots, z_n)$ have strictly positive elements, and let $Q = Q(z)$ be the $n \times n$ diagonal matrix whose i th diagonal element is z_i^{-1} . Also, define $\Omega = (X^\top Q^{-1} X)^{-1}$ and $\mu = (X^\top Q^{-1} X)^{-1} X^\top Q^{-1} y$. For each $s \in [0, \infty)$, define a univariate density as follows

$$\psi(u; s) = b(s) u^{d/2} e^{-su/2} h(u), \tag{2}$$

where $b(s)$ is the normalizing constant. The DA algorithm uses draws from the inverse Wishart (\mathcal{IW}_d) and matrix normal ($\mathcal{N}_{p,d}$) distributions. These densities are defined in the Appendix. If the current state of the DA Markov chain is $(\beta_m, \Sigma_m) = (\beta, \Sigma)$, then we simulate the new state, $(\beta_{m+1}, \Sigma_{m+1})$, using the following three-step procedure.

Iteration $m + 1$ of the DA algorithm:

Step 1: Draw Z_1, \dots, Z_n independently in such a way that, for each $i \in \{1, \dots, n\}$, $Z_i \sim \psi\{\cdot; (\beta^\top x_i - y_i)^\top \Sigma^{-1} (\beta^\top x_i - y_i)\}$, and call the result $z = (z_1, \dots, z_n)$.

Step 2: Draw $\Sigma_{m+1} \sim \mathcal{IW}_d\{n - p + 2a - d - 1, (y^\top Q^{-1} y - \mu^\top \Omega^{-1} \mu)^{-1}\}$.

Step 3: Draw $\beta_{m+1} \sim \mathcal{N}_{p,d}(\mu, \Omega, \Sigma_{m+1})$.

Download English Version:

<https://daneshyari.com/en/article/7546703>

Download Persian Version:

<https://daneshyari.com/article/7546703>

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