Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Geometric Ergodicity of Trans-Dimensional Markov Chain Monte Carlo Algorithms

Qian Qin

To cite this article: Qian Qin (24 Dec 2024): Geometric Ergodicity of Trans-Dimensional Markov Chain Monte Carlo Algorithms, Journal of the American Statistical Association, DOI: 10.1080/01621459.2024.2427432

To link to this article: https://doi.org/10.1080/01621459.2024.2427432

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2024.2427432

Geometric Ergodicity of Trans-Dimensional Markov Chain Monte Carlo Algorithms

Qian Qin

School of Statistics, University of Minnesota, Minneapolis, MN

ABSTRACT

This article studies the convergence properties of trans-dimensional MCMC algorithms when the total number of models is finite. It is shown that, for reversible and some nonreversible trans-dimensional Markov chains, under mild conditions, geometric convergence is guaranteed if the Markov chains associated with the within-model moves are geometrically ergodic. This result is proved in an L^2 framework using the technique of Markov chain decomposition. While the technique was previously developed for reversible chains, this work extends it to the point that it can be applied to some commonly used nonreversible chains. The theory herein is applied to reversible jump algorithms for three Bayesian models: a probit regression with variable selection, a Gaussian mixture model with unknown number of components, and an autoregression with Laplace errors and unknown model order. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received August 2023
Accepted November 2024

KEYWORDS
Convergence rate; Markov chain decomposition; Reversible jump; Spectral gap

1. Introduction

In many statistical setups, the parameter space of interest is a union of disjoint subsets, where each subset corresponds to a model, and the dimensions of the subsets need not be the same. Trans-dimensional Markov chain Monte Carlo (MCMC) is a class of algorithms for sampling from distributions defined on such spaces, which allows for model selection as well as parameter estimation. This type of algorithm, especially the reversible jump MCMC developed by Green (1995), has been applied to important problems like change-point estimation (Green 1995), autoregression models (Troughton and Godsill 1998; Ehlers and Brooks 2002; Vermaak et al. 2004), variable selection (Chevallier, Fearnhead, and Sutton 2023), wavelet models (Cornish and Littenberg 2015) etc. The current article aims to provide conditions on geometric ergodicity for trans-dimensional Markov chains when the total number of models is finite.

Let K be a finite set whose elements are referred to as "models." A model k in K is associated with a non-empty measurable space $(\mathsf{Z}_k, \mathcal{A}_k)$ and a nonzero finite measure Ψ_k on $(\mathsf{Z}_k, \mathcal{A}_k)$. Let $\mathsf{X} = \bigcup_{k\in\mathsf{K}}\{k\} \times \mathsf{Z}_k$, and let $\mathcal{A}$ be the sigma algebra generated by sets of the form $\{k\} \times \mathsf{A}$, where $k \in \mathsf{K}$ and $\mathsf{A} \in \mathcal{A}_k$. Consider the task of sampling from the probability measure Π on $(\mathsf{X}, \mathcal{A})$ such that

$$\Pi(\{k\} \times \mathsf{A}) = \frac{\Psi_k(\mathsf{A})}{\sum_{k'\in\mathsf{K}} \Psi_{k'}(\mathsf{Z}_{k'})}, \quad k \in \mathsf{K},\ \mathsf{A} \in \mathcal{A}_k. \tag{1}$$

Suppose that a procedure generates a random element $(K, Z) \sim \Pi$. Then, for $k \in \mathsf{K}$, $\Psi_k(\mathsf{Z}_k)/\sum_{k'\in\mathsf{K}} \Psi_{k'}(\mathsf{Z}_{k'})$ gives the probability of $K = k$, and $\Phi_k(\cdot) = \Psi_k(\cdot)/\Psi_k(\mathsf{Z}_k)$ gives the conditional distribution of Z given $K = k$.

In practice, Π is often intractable, prompting the use of trans-dimensional MCMC methods. A central goal of the current work is to provide verifiable sufficient conditions for trans-dimensional MCMC algorithms to be geometrically convergent in the L^2 distance. When $\mathcal{A}$ is countably generated and the Markov chain is φ-irreducible, L^2 geometric convergence implies the classical notion of Π-a.e. geometric ergdoicity. See Roberts and Rosenthal (1997), Roberts and Tweedie (2001), and Gallegos-Herrada, Ledvinka, and Rosenthal (2023). Geometric ergodicity is one of the key conditions ensuring the reliability of MCMC estimation. It guarantees a central limit theorem (CLT) for ergodic sums (Jones and Hobert 2001); moreover, consistent uncertainty assessment through asymptotically valid confidence intervals is possible under geometric ergodicity (Vats, Flegal, and Jones 2019).

The convergence behavior of trans-dimensional MCMC algorithms is in general far from well understood. Roberts and Rosenthal (2006) established some general conditions for trans-dimensional Markov chains to be Harris recurrent. Geometric ergodicity of some specific trans-dimensional algorithms was established in Geyer and Møller (1994), Andrieu and Doucet (1999), Ortner, Descombes, and Zerubia (2006), and Schreck et al. (2015). Existing proofs of geometric ergodicity often rely on drift and minorization conditions, or in some simple situations, Doeblin's condition. The current work instead uses the decomposition of Markov chains, a remarkable technique developed by Caracciolo, Pelissetto, and Sokal (1992) and documented in Madras and Randall (2002). This technique allows one to decompose the dynamic of a trans-dimensional Markov chain into within- and between-model movements, which can be analyzed separately. Using an extended version of this technique and exploiting the assumption that $|\mathsf{K}| < \infty$, Theorem 1 is established. This result describes a divide-and-conquer paradigm that enables one to establish geometric

CONTACT Qian Qin qqin@umn.edu School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church St SE, Minneapolis, MN 55455.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

convergence of the trans-dimensional chain by combining the geometric ergodicity of its within-model components. Quantitative bounds on the convergence rate will also be provided; see Theorem 7.

Previously, Markov chain decomposition has found its use in important problems like simulated and parallel tempering. See, for example, Madras and Randall (2002), Woodard, Schmidler, and Huber (2009), and Ge, Lee, and Risteski (2018). This technique can be used to analyze a Markov chain whose state space can be partitioned into subsets such that, within each subset, the Markov chain's behavior is easy to analyze. It was originally developed for reversible Markov chains. The current work provides an extended version of the technique, given in Lemma 3, that can deal with some important nonreversible chains. (Despite their name, reversible jump algorithms can often be nonreversible.)

The theory developed herein is applied to reversible jump MCMC algorithms for three practical Bayesian models: a probit regression model with variable selection, a Gaussian mixture model with unknown number of components, and an autoregressive model with Laplace errors and unknown model order. Among these algorithms, only one is assuredly reversible. For each algorithm, we demonstrate geometric convergence. Enabled by geometric ergodicity, we also conduct Monte Carlo uncertainty assessments.

Finally, it must be emphasized that verifying geometric ergodicity is but one of the first steps toward fully understanding the convergence behavior of an MCMC algorithm. A chain being geometrically convergent does not ensure that it has a fast convergence rate. While this work does provide a quantitative convergence rate bound, calculating the quantities involved in the bound can be practically challenging. Obtaining sharp estimates for the convergence rate remains an open problem for most practical trans-dimensional MCMC algorithms.

The rest of this article is organized as follows. Following a quick overview of the main qualitative result of this article, Section 2 contains some preliminary facts on the L^2 theory of Markov chains. The main technical results involving Markov chain decomposition and the convergence rate of trans-dimensional MCMC are given in Section 3. One toy and two practical examples are studied in Section 4, followed by a brief discussion in Section 5. Supplement I contains some minor results and technical proofs. Supplement II contains yet another practical example.

1.1. Conditions for Geometric Ergodicity: An Overview

Consider a trans-dimensional Markov chain $(X(t))_{t=0}^{\infty} = (K(t), Z(t))_{t=0}^{\infty}$ whose state space is X. Let $P : \mathsf{X} \times \mathcal{A} \to [0,1]$ be its Markov transition kernel (Mtk), that is, for $(k,z) \in \mathsf{X}$ and $\mathsf{A} \in \mathcal{A}$, $P((k,z), \mathsf{A})$ is understood as the conditional probability of $(K(t+1), Z(t+1)) \in \mathsf{A}$ given $(K(t), Z(t)) = (k,z)$. Suppose that Π is a stationary distribution of this chain, that is, $\Pi = \Pi P$, or more explicitly, for $\bigcup_{k\in\mathsf{K}}\{k\} \times \mathsf{A}_k \in \mathcal{A}$,

$$\sum_{k'\in\mathsf{K}} \Psi_{k'}(\mathsf{A}_{k'}) = \sum_{k\in\mathsf{K}}\sum_{k'\in\mathsf{K}} \int_{\mathsf{Z}_k} \Psi_k(\mathrm{d}z) P((k,z), \{k'\} \times \mathsf{A}_{k'}).$$

The main results of this article are stated in terms the L^2 theory for Markov chains, which is reviewed in Section 2. Essentially, if $\mathcal{T}(\cdot,\cdot)$ is an Mtk that has a stationary distribution ω, then $\mathcal{T}$ can be regarded as a bounded linear operator on a certain Hilbert space $L_0^2(\omega)$. A sufficient condition for the corresponding chain to be L^2 geometrically convergent is that the operator norm of some power of $\mathcal{T}$ is less than one.

One of the main results of this article is stated below. See Section 3 for more details.

Theorem 1. Assume that each of the following conditions holds for the trans-dimensional chain:

(H1) There exist a positive integer t_0 and a sequence of Mtks $P_k : \mathsf{Z}_k \times \mathcal{A}_k \to [0,1]$, $k \in \mathsf{K}$, such that the following properties hold for each k:

- (i) $\Phi_k P_k = \Phi_k$, where $\Phi_k(\cdot) = \Psi_k(\cdot)/\Psi_k(\mathsf{Z}_k)$ is the normalization of $\Psi_k(\cdot)$.
- (ii) When P_k is regarded as an operator on $L_0^2(\Phi_k)$, the norm of its t_0'th power $P_k^{t_0}$ is strictly less than 1.
- (iii) there exists a constant $c_k > 0$ such that $P((k,z), \{k\} \times \mathsf{A}) \geq c_k P_k(z, \mathsf{A})$ for $z \in \mathsf{Z}_k$ and $\mathsf{A} \in \mathcal{A}_k$.

(H2) The between-model movements are irreducible. To be precise, the Mtk $\bar{P} : \mathsf{K} \times 2^{\mathsf{K}} \to [0,1]$ given by $\bar{P}(k, \{k'\}) = \int_{\mathsf{Z}_k} \Phi_k(\mathrm{d}z) P((k,z), \{k'\} \times \mathsf{Z}_{k'})$, $k, k' \in \mathsf{K}$, is irreducible.

Then the norm of P^{t_0} is strictly less than one, and the trans-dimensional chain is $L^2(\Pi)$ geometrically convergent.

Note that $\bar{P}(k, \{k'\})$ can be understood as the average probability flow from model k to model k', and thus $\bar{P}$ characterizes the between movements. Evidently, (H2) holds as long as the chain is Π-irreducible. We give a rigorous proof of this assertion in Section A of Supplement I. Hence, in practice, (H2) usually trivially holds.

Trans-dimensional MCMC algorithms typically involve a within-model move type, where the underlying chain stays in a model, say k, with probability c_k, and move according to an Mtk P_k such that $\Phi_k P_k = \Phi_k$. Then such c_k and P_k satisfy (i) and (iii) in (H1). Condition (ii) in (H1) requires a careful analysis of the within-model moves of an algorithm. According to Lemma 2 below, in several important situations, this condition is implied by the geometric ergodicity of the chain associated with P_k, or some closely related Markov chain whose state space is Z_k. Since the space Z_k is typically of a fixed dimension, the hope is that chains that move in Z_k can be analyzed using well-established tools such as drift and minorization or functional inequalities.

Lemma 2. Let k be in K. Suppose that $\mathcal{A}_k$ is countably generated, $\Phi_k P_k = \Phi_k$, and the chain associated with P_k is φ-irreducible. Then, in each of the following situations, (ii) in (H1) holds with $t_0 = 1$.

- (i) P_k defines a Φ_k-a.e. geometrically ergodic chain that is reversible with respect to Φ_k.
- (ii) P_k defines a deterministic-scan Gibbs chain with two components that is Φ_k-a.e. geometrically ergodic.
- (iii) P_k defines a deterministic-scan Gibbs chain, and there exists a Φ_k-a.e. geometrically ergodic random-scan Gibbs chain based on the same set of conditional distributions.

Proof. For (i), see Theorem 2.1 of Roberts and Rosenthal (1997) and Theorem 2 of Roberts and Tweedie (2001). For (ii), see Lemma 3.2 and Proposition 3.5 of Qin and Jones (2022). For (iii), see Theorem 3.1 of Chlebicka, Łatuszyński, and Miasojedow (2024) and invoke (i). □

In Section 4.3, we give an example of establishing and utilizing (ii) in (H1) with $t_0 > 1$.

2. Preliminaries

Let $(\mathsf{Y}, \mathcal{F}, \omega)$ be a generic probability space. Let $L^2(\omega)$ be the Hilbert space of real functions $f : \mathsf{Y} \to \mathbb{R}$ that are square integrable with respect to ω, with the inner product between two functions defined as $\langle f, g\rangle_\omega = \int_{\mathsf{Y}} f(y)g(y)\,\omega(\mathrm{d}y)$, and the norm defined as $\|f\|_\omega = \sqrt{\langle f, f\rangle_\omega}$. Denote by $L^2_0(\omega)$ the subspace of $L^2(\omega)$ that consists of functions f such that $\omega f := \langle f, \mathbf{1}_{\mathsf{Y}}\rangle_\omega = 0$, where $\mathbf{1}_{\mathsf{Y}}(y) = 1$ for $y \in \mathsf{Y}$. A probability measure μ on $(\mathsf{Y}, \mathcal{F})$ is said to be in $L^2_*(\omega)$ if $\mathrm{d}\mu/\mathrm{d}\omega$ exists and is in $L^2(\omega)$. For two probability measures μ and ν in $L^2_*(\omega)$, define their L^2 distance by $\|\mu - \nu\|_\omega = \|\mathrm{d}\mu/\mathrm{d}\omega - \mathrm{d}\nu/\mathrm{d}\omega\|_\omega$.

Let $\mathcal{T} : \mathsf{Y} \times \mathcal{F} \to [0, 1]$ be an Mtk whose stationary distribution is ω. For a probability measure μ on $\mathcal{F}$, define $\mu\mathcal{T}^t(\cdot) = \int_{\mathsf{Y}} \mu(\mathrm{d}y)\mathcal{T}^t(y, \cdot)$, where $\mathcal{T}^t$ is the corresponding t-step Mtk. We say $\mathcal{T}$ is $L^2(\omega)$ geometrically convergent if there exist $\rho < 1$ and a function $C : L^2_*(\omega) \to [0, \infty)$ such that for $\mu \in L^2_*(\omega)$ and $t \geq 1$,

$$\|\mu\mathcal{T}^t - \omega\|_\omega \leq C(\mu)\rho^t. \tag{2}$$

Let $\|\cdot\|_{\mathrm{TV}}$ be the total variance distance between two probability measures. $\mathcal{T}$ is said to be ω-a.e. geometrically ergodic if there exist $\rho < 1$ and $C : \mathsf{Y} \to [0, \infty)$ such that, for ω-almost every $y \in \mathsf{Y}$ and $t \geq 1$, $\|\mathcal{T}^t(y, \cdot) - \omega(\cdot)\|_{\mathrm{TV}} \leq C(y)\rho^t$. Results from Roberts and Tweedie (2001) indicate that when $\mathcal{F}$ is countably generated, if the chain is $L^2(\omega)$ geometrically convergent, then it is ω-a.e. geometrically ergodic; the converse holds if $\mathcal{T}$ is reversible with respect to ω. See also Roberts and Rosenthal (1997) and Gallegos-Herrada, Ledvinka, and Rosenthal (2023).

The Mtk $\mathcal{T}$ can be understood as a linear operator on $L^2_0(\omega)$: for $f \in L^2_0(\omega)$, $\mathcal{T}f(\cdot) = \int_{\mathsf{Y}} \mathcal{T}(\cdot, \mathrm{d}y)f(y)$. One can use the Cauchy-Schwarz inequality to show that the L^2 norm of $\mathcal{T}$, defined as

$$\|\mathcal{T}\|_\omega = \sup_{f \in L^2_0(\omega)\setminus\{0\}} \frac{\|\mathcal{T}f\|_\omega}{\|f\|_\omega},$$

is no greater than 1. The operator norms of $\mathcal{T}$ and its powers quantify the Markov chain's convergence rate, with smaller norms indicating faster convergence. Indeed, if s is a positive integer, then (2) holds for all $\mu \in L^2_*(\omega)$ and $t \geq 1$ with $\rho = \|\mathcal{T}^s\|_\omega^{1/s}$ and some $C(\cdot)$. See Theorem 2.1 of Roberts and Rosenthal (1997) for more details on the interpretation of $\|\mathcal{T}\|_\omega$.

The bounded operator $\mathcal{T}$ has a unique adjoint $\mathcal{T}^*$. It is well-known that $\|\mathcal{T}\|^2_\omega = \|\mathcal{T}^*\|^2_\omega = \|\mathcal{T}\mathcal{T}^*\|_\omega = \|\mathcal{T}^*\mathcal{T}\|_\omega$. The Mtk $\mathcal{T}$ is reversible with respect to ω if and only if the operator $\mathcal{T}$ is self-adjoint, that is, $\mathcal{T} = \mathcal{T}^*$. The operator $\mathcal{T}$ is positive semidefinite if it is self-adjoint, and $\langle \mathcal{T}f, f\rangle_\omega \geq 0$ for $f \in L^2_0(\omega)$.

When $\mathcal{T}$ is self-adjoint, its spectral gap is defined to be

$$\mathrm{Gap}_\omega(\mathcal{T}) = 1 - \sup_{f \in L^2_0(\omega)\setminus\{0\}} \frac{\langle f, \mathcal{T}f\rangle_\omega}{\|f\|^2_\omega}.$$

Note that $\mathrm{Gap}_\omega(\mathcal{T}) \geq 1 - \|\mathcal{T}\|_\omega \geq 0$. If $\mathcal{T}$ is positive semidefinite, $\|\mathcal{T}\|_\omega = 1 - \mathrm{Gap}_\omega(\mathcal{T})$.

3. Convergence Analysis

3.1. Markov Chain Decomposition

This section describes the main probabilistic tool for proving Theorem 1.

Again, let $(\mathsf{Y}, \mathcal{F}, \omega)$ be a probability space. Suppose that Y can be partitioned into a collection of disjoint subsets, $(\mathsf{Y}_k)_{k\in\mathsf{K}}$. For this section, allow K to be countably infinite. Assume that $\omega(\mathsf{Y}_k) > 0$ for each k. Caracciolo, Pelissetto, and Sokal (1992) proposed a framework for analyzing a Markov chain moving in Y by decomposing its dynamic into local movements within a subset Y_k and global movements across the disjoint subsets. The key technical result, published in Madras and Randall (2002), is stated for reversible chains (see also, e.g., Guan and Krone 2007; Woodard, Schmidler, and Huber 2009). Here, it is extended to a possibly nonreversible setting.

For $k \in \mathsf{K}$, let $\mathcal{F}_k$ be the restriction of $\mathcal{F}$ on Y_k, and let $\omega_k(\mathsf{B}) = \omega(\mathsf{B})/\omega(\mathsf{Y}_k)$ for $\mathsf{B} \in \mathcal{F}_k$. Let $\bar{\omega}(\{k\}) = \omega(\mathsf{Y}_k)$ for $k \in \mathsf{K}$. For an Mtk $\mathcal{S} : \mathsf{Y} \times \mathcal{F} \to [0, 1]$ such that $\omega\mathcal{S} = \omega$, let $\bar{\mathcal{S}}$ be an Mtk on the discrete space K such that, for $k, k' \in \mathsf{K}$,

$$\bar{\mathcal{S}}(k, \{k'\}) = \frac{1}{\omega(\mathsf{Y}_k)}\langle \mathbf{1}_{\mathsf{Y}_k}, \mathcal{S}\mathbf{1}_{\mathsf{Y}_{k'}}\rangle_\omega = \frac{1}{\omega(\mathsf{Y}_k)}\int_{\mathsf{Y}_k} \omega(\mathrm{d}y)\,\mathcal{S}(y, \mathsf{Y}_{k'}).$$

Then $\bar{\omega}\bar{\mathcal{S}} = \bar{\omega}$. It can be checked that $\bar{\mathcal{S}}$ defines a self-adjoint (resp. positive semidefinite) operator on $L^2_0(\bar{\omega})$ whenever $\mathcal{S}$ is self-adjoint (resp. positive semidefinite). In the same vein, define the Mtk $\overline{\mathcal{S}^*\mathcal{S}}$ on K, which takes the form

$$\begin{aligned}\overline{\mathcal{S}^*\mathcal{S}}(k, \{k'\}) &= \frac{1}{\omega(\mathsf{Y}_k)}\langle \mathbf{1}_{\mathsf{Y}_k}, \mathcal{S}^*\mathcal{S}\mathbf{1}_{\mathsf{Y}_{k'}}\rangle_\omega \\ &= \frac{1}{\omega(\mathsf{Y}_k)}\int_{\mathsf{Y}} \omega(\mathrm{d}y)\,\mathcal{S}(y, \mathsf{Y}_k)\mathcal{S}(y, \mathsf{Y}_{k'}).\end{aligned}$$

As long as $\omega\mathcal{S} = \omega$, $\overline{\mathcal{S}^*\mathcal{S}}$ defines a positive semidefinite operator on $L^2_0(\bar{\omega})$.

Below is the key technical lemma of this section.

Lemma 3. Let $\mathcal{T}$ and $\mathcal{S}$ be Mtks such that $\omega\mathcal{T} = \omega\mathcal{S} = \omega$. Suppose that for $k \in \mathsf{K}$, there exists an Mtk $\mathcal{T}_k : \mathsf{Y}_k \times \mathcal{F}_k \to [0, 1]$ such that $\omega_k\mathcal{T}_k = \omega_k$. Assume further that there exists $c \in [0, 1]$ such that $\mathcal{T}(y, \mathsf{B}) \geq c\mathcal{T}_k(y, \mathsf{B})$ for $k \in \mathsf{K}$, $y \in \mathsf{Y}_k$ and $\mathsf{B} \in \mathcal{F}_k$. Then

$$1 - \|\mathcal{T}\mathcal{S}^*\|^2_\omega \geq c^2\left(1 - \sup_{k\in\mathsf{K}} \|\mathcal{T}_k\|^2_{\omega_k}\right)\mathrm{Gap}_{\bar{\omega}}(\overline{\mathcal{S}^*\mathcal{S}}). \tag{3}$$

In particular, if furthermore $\mathcal{S} = \mathcal{T}$, then

$$1 - \|\mathcal{T}\|^4_\omega \geq c^2\left(1 - \sup_{k\in\mathsf{K}} \|\mathcal{T}_k\|^2_{\omega_k}\right)\mathrm{Gap}_{\bar{\omega}}(\overline{\mathcal{T}^*\mathcal{T}}).$$

Remark 4. Lemma 3 extends Theorem A.1 of Madras and Randall (2002), originally formulated by Caracciolo, Pelissetto, and Sokal (1992). From Caracciolo, Pelissetto, and Sokal's (1992) result, it can be deduced that, if, in addition to the assumptions in Lemma 3, $\mathcal{T}$ and $\mathcal{S}$ are reversible with respect to ω, $\mathcal{T}_k$ is

reversible with respect to ω_k for $k \in \mathsf{K}$, and the operator $\mathcal{S}$ is positive semidefinite, then

$$\mathrm{Gap}_{\omega}(\mathcal{S}^{1/2}\mathcal{T}\mathcal{S}^{1/2}) \geq c \inf_{k\in\mathsf{K}} \mathrm{Gap}_{\omega_k}(\mathcal{T}_k)\mathrm{Gap}_{\bar{\omega}}(\bar{\mathcal{S}}).$$

In particular, if furthermore $\mathcal{S} = \mathcal{T}$, then

$$1 - \|\mathcal{T}\|_{\omega}^2 = \mathrm{Gap}_{\omega}(\mathcal{T}^2) \geq c \inf_{k\in\mathsf{K}} \mathrm{Gap}_{\omega_k}(\mathcal{T}_k)\mathrm{Gap}_{\bar{\omega}}(\bar{\mathcal{T}}).$$

Remark 5. The proof of Lemma 3 is given in Section B of Supplement I. It adopts the general idea of the proof of Theorem A.1 of Madras and Randall (2002), with alterations made to tackle non-reversibility. Notably, Madras and Randall (2002) studied how a reversible $\mathcal{T}$ acts on functions of the form $\mathcal{S}^{1/2}f$ for $f \in L_0^2(\omega)$ by decomposing the Dirichlet form $\|\mathcal{S}^{1/2}f\|_{\omega}^2 - \langle \mathcal{S}^{1/2}f, \mathcal{T}\mathcal{S}^{1/2}f\rangle_{\omega}$ into local and global components. Here, we study how a possibly nonreversible $\mathcal{T}$ acts on functions of the form $\mathcal{S}^*f$ by decomposing $\|\mathcal{S}^*f\|_{\omega}^2 - \|\mathcal{T}\mathcal{S}^*f\|_{\omega}^2$.

Remark 6. Using standard techniques, it is straightforward to derive the following bound in the opposite direction of those given in Lemma 3 and Madras and Randall (2002):

$$1 - \|\mathcal{T}\|_{\omega}^2 \leq \mathrm{Gap}_{\bar{\omega}}(\overline{\mathcal{T}^*\mathcal{T}}) \leq 1 - \|\bar{\mathcal{T}}\|_{\bar{\omega}}^2.$$

We provide a brief derivation at the end of Section B in Supplement I.

3.2. Geometric Convergence of the Trans-Dimensional Chain

Lemma 3 can be used to construct an upper bound on the norm of P^t for some $t \geq 1$, where P is the Mtk of the trans-dimensional chain defined in the Introduction.

Recall the definitions of $\bar{\mathcal{S}}$ and $\overline{\mathcal{S}^*\mathcal{S}}$, and consider letting $(\mathsf{Y}, \mathcal{F}, \omega) = (\mathsf{X}, \mathcal{A}, \Pi)$ and $\mathcal{S} = P$. Then $\bar{P}$ is defined as in (H2), and

$$\overline{P^*P}(k, \{k'\}) = \frac{1}{\Psi_k(\mathsf{Z}_k)} \sum_{k''\in\mathsf{K}} \int_{\mathsf{Z}_{k''}} \Psi_{k''}(\mathrm{d}z) P((k'', z), \{k\} \times \mathsf{Z}_k) \\ P((k'', z), \{k'\} \times \mathsf{Z}_{k'}).$$

These two Mtks describe the between-model movements of the trans-dimensional chain. $\bar{P}(k, \{k'\})$ can be understood as the average probability of moving from model k to model k'. $\overline{P^*P}(k, \{k'\})$ is similar, but with P replaced by P^*P. Indeed, under mild conditions, P^* and thus P^*P can be seen as Mtks that leave Π invariant (Paulin 2015; Choi 2020), and one can show that $\overline{P^*P}(k, \{k'\}) = \int_{\mathsf{Z}_k} \Phi_k(\mathrm{d}z) P^*P((k, z), \{k'\} \times \mathsf{Z}_{k'})$. P^*P is called the "multiplicative reversibilization" of P. Multiplicative reversibilizations are commonly investigated for nonreversible chains since at least Fill (1991). If P defines a self adjoint (resp. positive semipositive) operator on $L_0^2(\Pi)$, then $\bar{P}$ defines a self adjoint (resp. positive semipositive) operator on $L_0^2(\bar{\Pi})$, where $\bar{\Pi}(\{k\}) = \Psi_k(\mathsf{Z}_k)/\sum_{k'\in\mathsf{K}} \Psi_{k'}(\mathsf{Z}_{k'})$ for $k \in \mathsf{K}$. On the other hand, $\overline{P^*P}$ always defines a positive semidefinite operator on $L_0^2(\bar{\Pi})$.

We now provide a quantitative bound concerning the convergence rate of the trans-dimensional chain.

Theorem 7. Just for this theorem, allow $|\mathsf{K}|$ to be countably infinite. Suppose that, for each $k \in \mathsf{K}$, there exists an Mtk $P_k : \mathsf{Z}_k \times \mathcal{A}_k \to [0, 1]$ such that $\Phi_k P_k = \Phi_k$. Suppose further that, for $k \in \mathsf{K}$, there exists $c_k > 0$ such that $P((k, z), \{k\} \times \mathsf{A}) \geq c_k P_k(z, \mathsf{A})$ for $z \in \mathsf{Z}_k$ and $\mathsf{A} \in \mathcal{A}_k$. Then, for any positive integer t,

$$1 - \|P^t\|_{\Pi}^4 \geq \left(\inf_{k\in\mathsf{K}} c_k^t\right)^2 \left(1 - \sup_{k\in\mathsf{K}} \|P_k^t\|_{\Phi_k}^2\right) \mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P}). \quad (4)$$

If, furthermore, P defines a positive semidefinite operator on $L_0^2(\Pi)$ and P_k is reversible with respect to Φ_k for $k \in \mathsf{K}$, then there is the simpler bound

$$1 - \|P\|_{\Pi}^2 \geq \left(\inf_{k\in\mathsf{K}} c_k\right) \left[\inf_{k\in\mathsf{K}} \mathrm{Gap}_{\Phi_k}(P_k)\right] \mathrm{Gap}_{\bar{\Pi}}(\bar{P}). \quad (5)$$

Proof. We will establish (4) using Lemma 3; (5) can be established in a similar fashion using the original Markov chain decomposition result in Remark 4.

Fix a positive integer t. In Lemma 3, take $(\mathsf{Y}, \mathcal{F}, \omega) = (\mathsf{X}, \mathcal{A}, \Pi)$, $\mathcal{T} = P^t$ and $\mathcal{S} = P$. For $k \in \mathsf{K}$, let $\mathsf{Y}_k = \{k\} \times \mathsf{Z}_k$. Then $\mathcal{F}_k$ consists of sets of the form $\{k\} \times \mathsf{A}$, where $\mathsf{A} \in \mathcal{A}_k$, and $\omega_k(\{k\} \times \mathsf{A}) = \Phi_k(\mathsf{A})$ for $\mathsf{A} \in \mathcal{A}_k$. For $k \in \mathsf{K}, z \in \mathsf{Z}_k$, and $\mathsf{A} \in \mathcal{A}_k$, let $\mathcal{T}_k((k, z), \{k\} \times \mathsf{A}) = P_k^t(z, \mathsf{A})$. Since $\Phi_k P_k = \Phi_k$, it holds that $\omega_k \mathcal{T}_k = \omega_k$. Since $P((k, z), \{k\} \times \mathsf{A}) \geq c_k P_k(z, \mathsf{A})$ for $z \in \mathsf{Z}_k$ and $\mathsf{A} \in \mathcal{A}_k$, it holds that, for $(k, z) \in \mathsf{Y}_k$ and $\{k\} \times \mathsf{A} \in \mathcal{F}_k$,

$$\mathcal{T}((k, z), \{k\} \times \mathsf{A}) \geq c_k^t P_k^t(z, \mathsf{A}) \geq c\mathcal{T}_k((k, z), \{k\} \times \mathsf{A}),$$

where $c = \inf_{k\in\mathsf{K}} c_k^t$. Thus, the assumptions of Lemma 3 are satisfied.

The next step is identifying the objects in (3). Obviously, $\|\mathcal{T}\mathcal{S}^*\|_{\omega} = \|P^tP^*\|_{\Pi}$. Standard arguments show that $\|\mathcal{T}_k\|_{\omega_k} = \|P_k^t\|_{\Phi_k}$. The distribution $\bar{\omega}$ corresponds to $\bar{\Pi}$, while the Mtk $\overline{\mathcal{S}^*\mathcal{S}}$ is $\overline{P^*P}$. Then, by Lemma 3,

$$1 - \|P^tP^*\|_{\Pi}^2 \geq c^2 \left(1 - \sup_{k\in\mathsf{K}} \|P_k^t\|_{\Phi_k}^2\right) \mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P}).$$

Finally, note that

$$\|P^t\|_{\Pi}^2 = \|P^tP^{*t}\|_{\Pi} \leq \|P^tP^*\|_{\Pi}\|P^{*t-1}\|_{\Pi} \\ = \|P^tP^*\|_{\Pi}\|P^{t-1}\|_{\Pi} \leq \|P^tP^*\|_{\Pi}.$$

The desired result then follows. □

Theorem 7 connects the convergence properties of the trans-dimensional chain, quantified by $\|P\|_{\Pi}$, to the convergence properties of the within- and between-model movements, quantified by the $\|P_k\|_{\Phi_k}$'s and $\mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P})$ (or $\mathrm{Gap}_{\bar{\Pi}}(\bar{P})$), respectively. In Section 4.1, we use a toy example to test the sharpness of the bounds in Theorem 7, and investigate how the within- and between-model components may affect $\|P\|_{\Pi}$ and its bound.

Remark 8. By Remark 6, we also have $1 - \|P\|_{\Pi}^2 \leq \mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P})$, and if P is reversible, $1 - \|P\|_{\Pi} \leq \mathrm{Gap}_{\bar{\Pi}}(\bar{P})$. Thus, $\|P\|_{\Pi}$ is controlled by $\mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P})$ from above and below. In particular, combining (4) with the above yields

$$\frac{1}{4}\left(\inf_{k\in\mathsf{K}} c_k\right)^2 \left(1 - \sup_{k\in\mathsf{K}} \|P_k\|_{\Phi_k}^2\right) \leq \frac{1 - \|P\|_{\Pi}}{\mathrm{Gap}_{\bar{\Pi}}(\overline{P^*P})} \leq 1.$$

Similarly, if P is positive semidefinite and P_k is reversible for each k,

$$\frac{1}{2}\left(\inf_{k\in\mathsf{K}} c_k\right)\left[\inf_{k\in\mathsf{K}} \mathrm{Gap}_{\Phi_k}(P_k)\right] \le \frac{1-\|P\|_\Pi}{\mathrm{Gap}_{\bar\Pi}(\bar P)} \le 1.$$

Quantities such as $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P})$ and $\|P_k\|_{\Phi_k}$ may be difficult to compute in practice. However, when $|\mathsf{K}| < \infty$, Theorem 7 immediately yields Theorem 1, which is stated again below:

Theorem 1. Assume that (H1) and (H2) hold. Then $\|P^{t_0}\|_\Pi < 1$, and P is $L^2(\Pi)$ geometrically convergent.

Proof. By Theorem 7, it suffices to show that

$$\left(\min_{k\in\mathsf{K}} c_k^{t_0}\right)^2\left(1-\max_{k\in\mathsf{K}}\|P_k^{t_0}\|_{\Phi_k}^2\right)\mathrm{Gap}_{\bar\Pi}(\overline{P^*P}) > 0.$$

By (iii) in (H1), $\min_{k\in\mathsf{K}} c_k^{t_0} > 0$. By (ii) in (H1), $\max_{k\in\mathsf{K}} \|P_k^{t_0}\|_{\Phi_k}^2 < 1$.

It remains to show that $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P}) > 0$. Assume the opposite, that is, $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P}) = 0$. Because K is finite and $\overline{P^*P}$ is reversible, this implies that the largest eigenvalue of $\overline{P^*P}$ is 1. It then follows that $\overline{P^*P}$ is reducible (Hairer 2006, Theorem 3.11). By (iii) in (H1), for $k, k' \in \mathsf{K}$,

$$\begin{aligned}\overline{P^*P}(k,\{k'\}) &\ge \int_{\mathsf{Z}_k} \Phi_k(\mathrm{d}z)P((k,z),\{k\}\times\mathsf{Z}_k)P((k,z),\{k'\}\times\mathsf{Z}_{k'})\\ &\ge c_k\bar P(k,\{k'\}).\end{aligned}$$

So $\bar P$ must be reducible as well. But this contradicts with (H2). Hence, $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P}) > 0$. □

Theorem 1 will be used to establish geometric convergence for two practical examples in Section 4 and another one in Section 5.

4. Examples

This section contains a toy example and two practical problems concerning variable selection and mixture models, respectively. For a third practical example concerning autoregression, see Supplement II.

4.1. A Toy Chain

We first use a toy algorithm to test the sharpness of the quantitative bounds in Theorem 7.

Let $k_{\max}$ and n be positive integers. Let $\mathsf{K} = \{1,\dots,k_{\max}\}$. Consider a simple scenario where all the Ψ_k's are the same. To be specific, for $k = 1,\dots,k_{\max}$, let $\mathsf{Z}_k = \{1,\dots,n\}$, and let Ψ_k be the counting measure on Z_k. Then Π is the uniform distribution on $\mathsf{X} = \bigcup_{k\in\mathsf{K}}\{k\}\times\mathsf{Z}_k$.

We consider a type of MCMC algorithm targeting Π. Given the current state $(k,z)\in\mathsf{X}$, the algorithm either makes a local or a global move, each with probability 1/2. A local move sends the underlying Markov chain to (k,z') where $z'\in\mathsf{Z}_k$, and a global move sends the chain to (k',z) where $k'\in\mathsf{K}$. We consider three types of local moves and two types of global moves. The three types of local moves are "fast," "slow," and "varied." In a fast local move, the chain randomly and uniformly selects $z'\in\mathsf{Z}_k$ to move to. In a slow local move, the chain can either stay in place with probability 1/2, or move to one of the within-model neighbors. Within a model k, two states z and z' are neighbors if they differ by 1, or if one of them is 1 and the other is n. In a varied local move, the movement of the chain varies with the current model k. To be precise, the chain stays in place with probability $1-1/k$, and move randomly and uniformly across Z_k with probability $1/k$. The two types of global moves are "fast" and "slow." In a fast global move, the chain randomly and uniformly selects a new model $k'\in\mathsf{K}$ to move to. In a slow global move, the chain can stay in place with probability 1/2, or move to one of the neighboring models. Two models k and k' are neighbors if they differ by 1, or if one of them is 1 and the other is $k_{\max}$. Each combination of local and global move types gives rise to a concrete algorithm, and we may define six algorithms in this manner.

For a given algorithm, let P be the Mtk of the underlying chain, and let P_k be the Mtk associated with the local movement within model k. One can check that, for each of the six algorithms, P is positive semidefinite, and so are the P_k's. The local and global behavior of P is summarized as follows.

- When the local move type is fast, $\|P_k\|_{\Phi_k} = 0$. When the local move type is slow, $\|P_k\|_{\Phi_k}$ is a function of n, and it goes to 1 as $n\to\infty$. When the local move is varied, $\|P_k\|_{\Phi_k} = 1-1/k$.
- When the global move type is fast, $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P}) = 3/4$ and $\mathrm{Gap}_{\bar\Pi}(\bar P) = 1/2$. When the local move type is slow, $\mathrm{Gap}_{\bar\Pi}(\overline{P^*P})$ is a function of $k_{\max}$, and it goes to 0 as $k_{\max}\to\infty$; the same goes for $\mathrm{Gap}_{\bar\Pi}(\bar P)$.

Using Theorem 7, we may construct upper bounds on $\|P\|_\Pi$. The bound derived from (4) with $c_k = 1/2$ and $t = 1$ will be denoted by $\|P\|_\Pi^\dagger$, while the bound derived from (5), which exploits reversibility, is denoted by $\|P\|_\Pi^\ddagger$.

Since X is finite, the true value of $\|P\|_\Pi$ can be computed. We test the performance of the bounds through numerical simulation, and the results are given in Figure 1. Table 1 loosely summarizes how the local and global move types affect the sharpness of $\|P\|_\Pi^\dagger$. The bound $\|P\|_\Pi^\ddagger$ is sharper than $\|P\|_\Pi^\dagger$, but the two bounds are comparable.

4.2. Variable Selection in Bayesian Probit Regression

4.2.1. The Model

For $i = 1,\dots,n$, let $x_i = (x_{i,1},\dots,x_{i,r})^\top \in \mathbb{R}^r$ be a known vector of predictors. Let $Y_1,\dots,Y_n$ be independent binary responses, where Y_i follows a Bernoulli distribution with success probability $F(A + x_i^\top B)$, with $F(\cdot)$ being the cumulative distribution function of the standard normal distribution. The scalar $A\in\mathbb{R}$ is an unknown intercept, while the vector $B = (B_1,\dots,B_r)^\top\in\mathbb{R}^r$ is an unknown regression coefficient.

To perform Bayesian variable selection, put a spike and slab prior on B. To be specific, let $K = (K_1,\dots,K_r)^\top \in \{0,1\}^r$. Let $\mathsf{J}_K = \{j\in\{1,\dots,r\} : K_j = 1\}$. Place a prior distribution on K that has probability mass function proportional to $p^{|\mathsf{J}_k|}$, where $p\in(0,1)$ is a hyperparameter. Assume that given K, the B_j's are independent. If $K_j = 0$, B_j is set to be zero; otherwise, B_j follows the $\mathrm{N}(0,\sigma^2)$ distribution, where σ

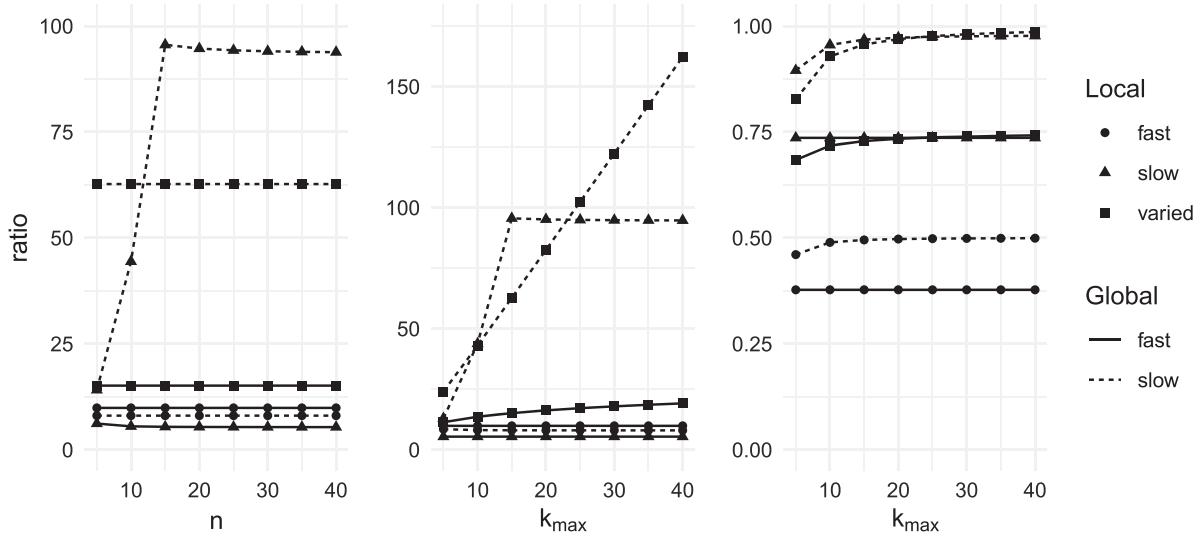

Figure 1. Performance of the quantitative bounds for the toy chain. Left: $(1-\|P\|_\Pi)/(1-\|P\|_\Pi^\dagger)$ for various values of n when $k_{\max}=15$. Middle: $(1-\|P\|_\Pi)/(1-\|P\|_\Pi^\dagger)$ for various values of $k_{\max}$ when $n=15$. Right: $(1-\|P\|_\Pi^\dagger)/(1-\|P\|_\Pi^\ddagger)$ for various values of $k_{\max}$ when $n=15$.

Table 1. Performance of the bound (4) for the toy chain.

	Locally fast	Locally slow	Locally varied
Globally fast	Well behaved	Well behaved	Deteriorates slowly as $k_{\max}\to\infty$
Globally slow	Well behaved	Situation dependent	Deteriorates as $k_{\max}\to\infty$

is a positive hyperparameter. Thus, K indicates the collection of relevant predictors. The intercept A is independent of (B,K) and follows the $\mathrm{N}(0,\sigma^2)$ distribution. Let $Z=(A,B^\top)^\top$. Having observed $Y=(Y_1,\ldots,Y_n)^\top$, the goal is to make inference about (K,Z).

The parameter space, that is, the range of (K,Z), is $\mathsf{X}=\bigcup_{k\in\mathsf{K}}\{k\}\times\mathsf{Z}_k$, where $\mathsf{K}=\{0,1\}^r$, and, for $k=(k_1,\ldots,k_r)\in\mathsf{K}$, Z_k is the set of $(\alpha,\beta_1,\ldots,\beta_r)\in\mathbb{R}^{r+1}$ such that $\beta_j=0$ whenever $k_j=0$. The posterior distribution Π of (K,Z) given $Y=y=(y_1,\ldots,y_n)$ is of the form (1). Evaluated at $z=(\alpha,\beta^\top)^\top\in\mathsf{Z}_k$, the density function of Ψ_k is

$$\pi(k,z\mid y)=\frac{1}{\sqrt{2\pi}\sigma}\left(\frac{p}{\sqrt{2\pi}\sigma}\right)^{|\mathsf{J}_k|}\exp\left(-\frac{\alpha^2+\sum_{j\in\mathsf{J}_k}\beta_j^2}{2\sigma^2}\right)\prod_{i=1}^n F\left(\alpha+\sum_{j\in\mathsf{J}_k}x_{i,j}\beta_j\right)^{y_i}\left[1-F\left(\alpha+\sum_{j\in\mathsf{J}_k}x_{i,j}\beta_j\right)\right]^{1-y_i}.$$

4.2.2. A Reversible Jump MCMC Algorithm

For a fixed model $k=(k_1,\ldots,k_r)$, one can use a well-known data augmentation algorithm (a type of reversible MCMC algorithm) devised by Albert and Chib (1993) to sample from Φ_k, the normalization of Ψ_k. The algorithm is now briefly described. Suppose that $\mathsf{J}_k=\{j_1,\ldots,j_d\}$, where $d=|\mathsf{J}_k|$ and $j_1<\cdots<j_d$. Let $M(k)$ denote the $n\times(d+1)$ matrix whose ith row is $(1,x_{i,j_1},\ldots,x_{i,j_d})^\top$. Given the current state $z=(\alpha,\beta_1,\ldots,\beta_r)^\top\in\mathsf{Z}_k$, the next state $z'=(\alpha',\beta_1',\ldots,\beta_r')^\top$ is drawn through the following procedure: Independently for $i=1,\ldots,n$, draw u_i from the $\mathrm{N}(\alpha+\sum_{j=1}^r x_{i,j}\beta_j,1^2)$ distribution truncated to $(0,\infty)$ if $y_i=1$ and to $(-\infty,0)$ otherwise. Let $u=(u_1,\ldots,u_n)^\top$. Draw $(\alpha',\beta_{j_1}',\ldots,\beta_{j_d}')^\top$ from the normal distribution

$$\mathrm{N}_{d+1}\left(\left[M(k)^\top M(k)+\sigma^{-2}I_{d+1}\right]^{-1}M(k)^\top u,\left[M(k)^\top M(k)+\sigma^{-2}I_{d+1}\right]^{-1}\right),$$

and set the remaining elements of z' to zero.

The reversible jump MCMC algorithm considered herein is a combination of the data augmentation algorithm and a standard reversible jump scheme. Given the current state (k,z), the algorithm randomly performs a U (update), B (birth), or D (death) move. It is assumed that the probability of choosing a move depends on (k,z) only through k, and the probabilities are denoted by $q_{\mathrm{U}}(k)$, $q_{\mathrm{B}}(k)$, and $q_{\mathrm{D}}(k)$, respectively.

- *U move*: Draw z' using one iteration of Albert and Chib's data augmentation algorithm. Set the new state to (k,z').
- *B move*: Randomly and uniformly choose an index j from J_k^c, where the complement is taken with respect to the set $\{1,\ldots,r\}$. Change the jth element of k to 1 and call the resultant binary vector k'. Draw b_* from some distribution on $\mathbb{R}$ associated with a density function $g(\cdot\mid k,k',z,y)$. Replace the $(j+1)$th element of $z=(\alpha,\beta_1,\ldots,\beta_r)^\top$ by b_*, and call the resultant vector z'. With probability

$$\min\left\{1,\frac{\pi(k',z'\mid y)\,q_{\mathrm{D}}(k')\,(|\mathsf{J}_k|+1)^{-1}}{\pi(k,z\mid y)\,q_{\mathrm{B}}(k)\,(r-|\mathsf{J}_k|)^{-1}g(b_*\mid k,k',z,y)}\right\},$$

set the next state to (k',z'); otherwise, keep the old state. This move type is available only when $|\mathsf{J}_k|<r$.

- *D move*: Randomly and uniformly choose an index j from J_k. Change the jth element of k to 0 and call the resultant binary vector k'. Let β_j be the $(j+1)$th element of z. Let z' be the vector obtained by changing the $(j+1)$th element of z to 0. With probability

$$\min\left\{1, \frac{\pi(k', z' \mid y)\, q_{\mathrm{B}}(k')\, (r - |\mathsf{J}_k| + 1)^{-1} g(\beta_j \mid k', k, z', y)}{\pi(k, z \mid y)\, q_{\mathrm{D}}(k)\, |\mathsf{J}_k|^{-1}}\right\},$$

set the next state to (k', z'); otherwise, keep the old state. This move type is available only when $|\mathsf{J}_k| > 0$.

The resultant trans-dimensional chain is reversible with respect to Π.

4.2.3. Convergence Analysis

For $k \in \mathsf{K}$, let P_k be the Mtk of the data augmentation chain targeting Φ_k. Chakraborty and Khare (2017) proved the following result regarding P_k using the drift and minorization technique. See also Roy and Hobert (2007).

Lemma 9. For $k \in \mathsf{K}$, P_k is Φ_k-a.e. geometrically ergodic.

Given k, the data augmentation Mtk P_k is reversible with respect to Φ_k. One can then establish geometric convergence for the reversible jump chain.

Proposition 10. Suppose that $q_{\mathrm{U}}(k) > 0$ for $k \in \mathsf{K}$, $q_{\mathrm{B}}(k) > 0$ when $|\mathsf{J}_k| < r$, $q_{\mathrm{D}}(k) > 0$ when $|\mathsf{J}_k| > 0$. Then the reversible jump chain is $L^2(\Pi)$-geometrically convergent and Π-a.e. geometrically ergodic.

Proof. Apply Theorem 1. Let P be the Mtk of the reversible jump chain. By Lemma 9 and (i) of Lemma 2, $\|P_k\|_{\Phi_k} < 1$ for $k \in \mathsf{K}$. It follows that, for $k \in \mathsf{K}$, (i) and (ii) of (H1) hold with $t_0 = 1$. Evidently, (iii) of (H1) also holds with $c_k = q_{\mathrm{U}}(k)$.

To verify (H2), note that for $k, k' \in \mathsf{K}$, $\bar{P}(k, \{k'\}) > 0$ whenever k and k' differ by at most 1 element. Then it is clear that $\bar{P}$ is irreducible. (Alternatively, note that P is Π-irreducible.)

The chain is thus $L^2(\Pi)$ geometrically convergent. By Theorem 1 of Roberts and Tweedie (2001), it is Π-a.e. geometrically ergodic. □

4.2.4. Application to a Dataset

Geometric ergodicity allows one to estimate the importance of features in a variable selection problem with confidence. The reversible jump algorithm is applied to the `Spambase` dataset (Hopkins et al. 1999). This dataset contains $n = 4601$ e-mails. The response Y_i indicates whether the ith e-mail is spam. Each e-mail is associated with $r = 57$ attributes, including the frequency of certain words and the length of sequences of consecutive capital letters. To perform variable selection, a spike and slab prior with $p = 0.5$ is used.

In the B move of the reversible jump algorithm, $g(\cdot \mid k, k', z, y)$ is chosen to be the density of a normal distribution, whose parameters are selected using ideas from Brooks, Giudici, and Roberts (2003). The probabilities of proposing birth and death moves are as follows:

$$q_{\mathrm{B}}(k) = \frac{1}{3}\min\left\{1, \frac{p\,(r - |\mathsf{J}_k|)}{|\mathsf{J}_k| + 1}\right\},$$

$$q_{\mathrm{D}}(k) = \frac{1}{3}\min\left\{1, \frac{|\mathsf{J}_k|}{p\,(r - |\mathsf{J}_k| + 1)}\right\}.$$

By Proposition 10, the reversible jump chain is Π-a.e. geometrically ergodic.

A chain $(K(t), Z(t))_{t=1}^m$ of length $m = 10^5$ is simulated. The quantities of interest are the posterior probabilities of $K_j = 1$ for $j = 1, \ldots, 57$, that is, the posterior probability of any given predictor being present in the regression model. They are estimated using the sample proportions $m^{-1}\sum_{t=1}^m K_j(t)$. Under geometric ergodicity, the sample proportions are asymptotically normally distributed, and the asymptotic variances can be consistently estimated using the batch means method (Jones et al. 2006). To avoid underestimation, which is a problem exhibited by batch means estimators when the Monte Carlo sample size is not sufficiently large (Flegal and Jones 2010, sec. 4), we add $(\log m)\sqrt{1/b_m^2 + b_m/m}$ to the estimated asymptotic variances, where $b_m \approx m^{0.6}$ is the batch size, and $1/b_m^2 + b_m/m$ is on the same order as the mean squared error of the batch means estimator (Flegal and Jones 2010, sec. 3). This adjustment is further discussed in Section E of Supplement I. We construct 95% simultaneous Wald confidence intervals for the posterior probabilities. Bonferroni correction is used here, although more sophisticated multivariate methods could be considered; see Section 5. The confidence intervals for the last 10 variables (attributes) are presented in Figure 2. This figure shows how important each predictor is according to the MCMC simulation, as well as the errors of their estimated importance.

4.3. Gaussian Mixture Model

4.3.1. The Model

Let $Y_1, \ldots, Y_n$ be iid random variables drawn from the mixture of K normal distributions. For $j = 1, \ldots, K$, let W_j be the weight associated with the jth normal distribution, and let U_j and T_j be, respectively, the mean and variance of that normal distribution. Equivalently, we may formulate the model as follows. Let $(Y_1, A_1), \ldots, (Y_n, A_n)$ be iid random vectors that take values in $\mathbb{R} \times \{1, \ldots, k\}$, where, for each i, Y_i given A_i follows the $\mathrm{N}(U_{A_i}, T_{A_i})$ distribution, and marginally $P(A_i = j) = W_j$ for $j = 1, \ldots, K$. Suppose that $Y = (Y_1, \ldots, Y_n)$ is observable, while $A = (A_1, \ldots, A_n)$, $W = (W_1, \ldots, W_K)$, $U = (U_1, \ldots, U_K)$, $T = (T_1, \ldots, T_K)$, and K are unknown.

To perform Bayesian analysis, we assume that K has a prior probability mass function $k \mapsto f_K(k)$ that is supported on a finite set $\mathsf{K} = \{1, \ldots, k_{\max}\}$. We then put a Dirichlet prior on W, inverse gamma priors on T, and normal priors on U. To address the label switching problem and enforce identifiability, we shall assume that $U_1 \le U_2 \le \cdots \le U_K$ in the prior distribution. To be precise, we shall assume that, given $K = k$, the prior density function of (W, T, U) evaluated at $(w, \tau, u) = ((w_1, \ldots, w_k), (\tau_1, \ldots, \tau_k), (u_1, \ldots, u_k))$ has the following form:

$$f_{W,T,U}(w, \tau, u \mid k) = \left\{\prod_{j=1}^k w_j^{\gamma-1} \frac{b^c}{\Gamma(c)} \tau_j^{-c-1} e^{-b/\tau_j} \frac{1}{\sqrt{2\pi\,\tau_0 \tau_j}}\right.$$

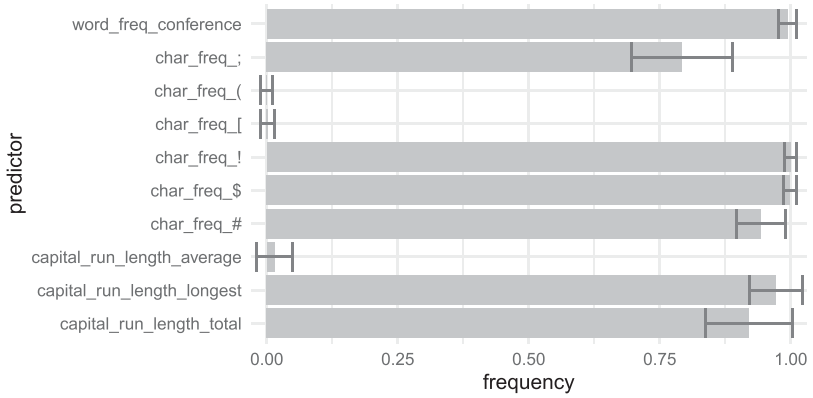

Figure 2. Estimated posterior probabilities of selecting each variable (attribute), with 95% simultaneous confidence intervals. Only the last 10 variables are shown.

$$\exp\left[-\frac{(u_j-u_0)^2}{2\tau_0\tau_j}\right]\mathbf{1}_{(0,\infty)}(\tau_j)\Bigg\}\frac{\Gamma(k\gamma)}{\Gamma(\gamma)^k}\mathbf{1}_{\mathsf{S}_k}(w)\,k!\,\mathbf{1}_{\mathsf{G}_k}(u). \tag{6}$$

In the above display, $\gamma, b, c, \tau_0, u_0$ are positive hyperparameters, $\Gamma(\cdot)$ is the gamma function, $\mathsf{S}_k = \{(w_1,\ldots,w_k) \in (0,1)^k : \sum_{j=1}^k w_j = 1\}$, $\mathsf{G}_k = \{(u_1,\ldots,u_k) \in \mathbb{R}^k : u_1 \le \cdots \le u_k\}$.

The un-normalized posterior density of (K, A, W, T, U) is then

$$\pi(k,\alpha,w,\tau,u \mid y) = \left\{\prod_{i=1}^n w_{\alpha_i}\frac{1}{\sqrt{2\pi\,\tau_{\alpha_i}}}\exp\left[-\frac{1}{2\tau_{\alpha_i}}(y_i-u_{\alpha_i})^2\right]\right\} f_{W,T,U}(w,\tau,u \mid k) f_K(k),$$

where α_i denotes the ith element of α. The corresponding measure Π has the form (1), with the density of Ψ_k given by $(\alpha, w, \tau, u) \mapsto \pi(k, \alpha, w, \tau, u \mid y)$. In this context, $\mathsf{Z}_k = \{1,\ldots,k\}^n \times \mathsf{S}_k \times (0,\infty)^k \times \mathsf{G}_k$.

4.3.2. A Reversible Jump Algorithm

The algorithm we consider is modified from Richardson and Green's (1997) reversible jump algorithm, with a new within-model move type and simplified between-model move types.

We shall first propose an algorithm for sampling from Φ_K when K is known. For $k \in \mathsf{K}$ and $(w,\tau,u) \in \mathsf{S}_k \times (0,\infty)^k \times \mathbb{R}^k$, let $\tilde{f}_{W,T,U}(w,\tau,u \mid k)$ be the same as (6) but without the constraint $\mathbf{1}_{\mathsf{G}_k}(u)$. It can be shown that, given $k \in \mathsf{K}$, for $\alpha \in \{1,\ldots,k\}^n$,

$$\begin{aligned}
&\tilde{\pi}_A(\alpha \mid k, y)\\
&:= \int_{\mathsf{S}_k\times(0,\infty)^p\times\mathbb{R}^p}\left\{\prod_{i=1}^n w_{\alpha_i}\frac{1}{\sqrt{2\pi\,\tau_{\alpha_i}}}\exp\left[-\frac{1}{2\tau_{\alpha_i}}(y_i-u_{\alpha_i})^2\right]\right\}\tilde{f}_{W,T,U}(w,\tau,u)\,\mathrm{d}(w,\tau,\mu)\\
&\propto \prod_{j=1}^k \frac{[\tau_0\,n_j(\alpha)+1]^{-1/2}\,\Gamma(n_j(\alpha)+\gamma)\,\Gamma(n_j(\alpha)/2+c)}{\{ss_j(\alpha)/2+u_0^2/(2\tau_0)-[s_j(\alpha)+u_0/\tau_0]^2/[2(n_j(\alpha)+1/\tau_0)]+b\}^{n_j(\alpha)/2+c}},
\end{aligned}$$

where $n_j(\alpha) = \sum_{i=1}^n \mathbf{1}_{\{j\}}(\alpha_i)$, $s_j(\alpha) = \sum_{i=1}^n y_i \mathbf{1}_{\{j\}}(\alpha_i)$, and $ss_j(\alpha) = \sum_{i=1}^n y_i^2 \mathbf{1}_{\{j\}}(\alpha_i)$. Let $z = (\alpha, w, \tau, u) \in \mathsf{Z}_k$ be the current state. The next state $z' = (\alpha', w', \tau', u')$ is drawn via the following steps:

1. Let $\alpha^{(0)} = (\sigma(\alpha_1),\ldots,\sigma(\alpha_n))$, where σ is a randomly and uniformly selected permutation of $\{1,\ldots,k\}$.
2. For i from 1 to n, do the following. Randomly and uniformly draw j_i from $\{1,\ldots,k\}$. Let $\alpha_{i\leftarrow j_i}^{(i-1)} \in \{1,\ldots,k\}^n$ be the vector obtained from $\alpha^{(i-1)}$ by changing its ith element to j_i. With probability
$$\min\left\{1, \frac{\tilde{\pi}_A(\alpha_{i\leftarrow j_i}^{(i-1)} \mid k, y)}{\tilde{\pi}_A(\alpha^{(i-1)} \mid k, y)}\right\},$$
let $\alpha^{(i)} = \alpha_{i\leftarrow j_i}^{(i-1)}$; otherwise, let $\alpha^{(i)} = \alpha^{(i-1)}$. Denote $\alpha^{(n)}$ by α''.
3. Draw $w'' = (w_1'',\ldots,w_k'')$ from the Dirichlet distribution with concentration parameter $(n_1(\alpha'')+\gamma,\ldots,n_k(\alpha'')+\gamma)$.
4. For $j = 1,\ldots,k$, independently draw τ_j'' from the inverse gamma distribution with shape parameter $c + n_j(\alpha'')/2$ and scale parameter
$$\frac{ss_j(\alpha'')}{2} + \frac{u_0^2}{2\tau_0} - \frac{[s_j(\alpha'')+\mu_0/\tau_0]^2}{2[n_j(\alpha'')+1/\tau_0]} + b.$$
5. For $j = 1,\ldots,k$, independently draw $u_1'',\ldots,u_k''$ from the normal distribution
$$\mathrm{N}\left(\frac{s_j(\alpha'')+u_0/\tau_0}{n_j(\alpha'')+1/\tau_0}, \frac{\tau_j''}{n_j(\alpha'')+1/\tau_0}\right).$$
6. Order $u_1'',\ldots,u_k''$ so that we find distinct indices $j_1,\ldots,j_k$ such that $u_{j_1}'' \le \cdots \le u_{j_k}''$. For $i = 1,\ldots,n$, find the index $\ell_i \in \{1,\ldots,k\}$ such that $\alpha_i'' = j_{\ell_i}$, and let $\alpha_i' = \ell_i$. For $\ell = 1,\ldots,k$, let $w_\ell' = w_{j_\ell}''$, $\tau_\ell' = \tau_{j_\ell}''$, and $u_\ell' = u_{j_\ell}''$.

We shall call the sampler the Metropolis re-ordering algorithm. The following lemma is proved in Section C of Supplement I.

Lemma 11. The underlying Markov chain of the Metropolis re-ordering algorithm leaves Φ_k invariant for $k \in \mathsf{K}$.

A nice property of the Metropolis re-ordering algorithm is that starting from any allocation $\alpha \in \{1,\dots,k\}^n$, it is possible for the chain to move to any allocation $\alpha' \in \{1,\dots,k\}^n$ in a single iteration.

Consider now a reversible jump algorithm for sampling from Π, which is part of an algorithm constructed by Richardson and Green (1997). When the current state is $(k,z) = (k,\alpha,w,\tau,u)$, the algorithm randomly performs a U, B, or D move. It is assumed that the probabilities of choosing these moves depend on the current state only through k, and are, respectively, $q_{\mathrm{U}}(k)$, $q_{\mathrm{B}}(k)$, $q_{\mathrm{D}}(k)$. The three move types are defined as follows:

- *U move*: Draw z' using one iteration of the Metropolis re-ordering algorithm. Set the new state to (k,z').
- *B move*: Draw (w_*,τ_*,u_*) from some distribution on $(0,1)\times(0,\infty)\times\mathbb{R}$ with density function $g(\cdot \mid k,\alpha,w,\tau,u)$. Find $\ell \in \{1,\dots,k+1\}$ such that $u_j \le u_*$ whenever $j<\ell$, and $u_j \ge u_*$ whenever $j \ge \ell$ — that is, u_* is the ℓth smallest number in the set $\{u_1,\dots,u_k,u_*\}$. Let $\alpha' = (\alpha'_1,\dots,\alpha'_n)$ be such that, for $i \in \{1,\dots,n\}$, $\alpha'_i = \alpha_i$ if $\alpha_i < \ell$ and $\alpha'_i = \alpha_i + 1$ if $\alpha_i \ge \ell$. Let
$$\begin{aligned} w' &= ((1-w_*)w_1,\dots,(1-w_*)w_{\ell-1},w_*,\\ &\quad (1-w_*)w_\ell,\dots,(1-w_*)w_k),\\ \tau' &= (\tau_1,\dots,\tau_{\ell-1},\tau_*,\tau_\ell,\dots,\tau_k),\\ u' &= (u_1,\dots,u_{\ell-1},u_*,u_\ell,\dots,u_k). \end{aligned}$$
With probability
$$\min\left\{1,\frac{\pi(k+1,\alpha',w',\tau',u'\mid y)\,q_{\mathrm{D}}(k+1)\,[\sum_{j=1}^{k+1}\mathbf{1}_{\{0\}}(n_j(\alpha'))]^{-1}\,(1-w_*)^{k-1}}{\pi(k,\alpha,w,\tau,u\mid y)\,q_{\mathrm{B}}(k)\,g(w_*,\tau_*,u_*\mid k,\alpha,w,\tau,u)}\right\},$$
set the new state to $(k+1,\alpha',w',\tau',u')$; otherwise, keep the old state. This move is available only when $k < k_{\max}$.
- *D move*: Let $\mathsf{E}_k(\alpha) \subset \{1,\dots,k\}$ be the set of j's such that $n_j(\alpha) = 0$. Keep the old state if $\mathsf{E}_k(\alpha) = \emptyset$, and follow the procedure below otherwise. Randomly and uniformly select ℓ from $\mathsf{E}_k(\alpha)$. Let $\alpha' = (\alpha'_1,\dots,\alpha'_n)$ be such that, for $i = 1,\dots,n$, $\alpha'_i = \alpha_i$ if $\alpha_i < \ell$ and $\alpha'_i = \alpha_i - 1$ if $\alpha_i > \ell$. Let
$$\begin{aligned} w' &= (w_1/(1-w_\ell),\dots,w_{\ell-1}/(1-w_\ell),\\ &\quad w_{\ell+1}/(1-w_\ell),w_k/(1-w_\ell)),\\ \tau' &= (\tau_1,\dots,\tau_{\ell-1},\tau_{\ell+1},\dots,\tau_k),\\ u' &= (u_1,\dots,u_{\ell-1},u_{\ell+1},\dots,u_k). \end{aligned}$$
With probability
$$\min\left\{1,\frac{\pi(k-1,\alpha',w',\tau',u'\mid y)\,q_{\mathrm{B}}(k-1)\,g(w_\ell,\tau_\ell,u_\ell\mid k-1,\alpha',w',\tau',u')}{\pi(k,\alpha,w,\tau,u\mid y)\,q_{\mathrm{D}}(k)\,[\sum_{j=1}^{k}\mathbf{1}_{\{0\}}(n_j(\alpha))]^{-1}\,(1-w_\ell)^{k-2}}\right\},$$
set the new state to $(k-1,\alpha',w',\tau',u')$; otherwise, keep the old state. This move is available only when $k > 1$.

The resultant chain has Π as its stationary distribution due to the reversible jump construction. On the other hand, the Metropolis-reordering algorithm is not classified as a well-known reversible algorithm, and it remains unclear whether the U move type is reversible or positive semidefinite.

For illustration, the algorithm is applied to the galaxy dataset described by Roeder (1990) and studied by Richardson and Green (1997). In the B move, w_* is drawn from the beta distribution with parameters 1 and n, τ_* is drawn from the inverse gamma distribution with shape parameter c and scale parameter b, and u_* is drawn from the $\mathrm{N}(0,\tau_0\tau_*)$ distribution. Following Green (1995), the birth and death probabilities are set to
$$q_{\mathrm{B}}(k) = \frac{1}{3}\min\left\{1,\frac{f_K(k+1)}{f_K(k)}\right\},$$
$$q_{\mathrm{D}}(k) = \frac{1}{3}\min\left\{1,\frac{f_K(k-1)}{f_K(k)}\right\}.$$
The trans-dimensional chain $(K(t),A(t),W(t),T(t),U(t))_{t=1}^{\infty}$ is simulated for 10^5 iterations. The empirical performance of the algorithm is shown in Figure 3. The predictive density evaluated at a point x is the sample average of $\sum_{j=1}^{K(t)} W_j(t) f(x \mid U_j(t),T_j(t))$, where $f(\cdot \mid u,\tau)$ denotes the density of the $\mathrm{N}(u,\tau)$ distribution. We can see that the sampler appears to perform well empirically, especially in terms of within-model moves.

4.3.3. Convergence Analysis

For $k \in \mathsf{K}$, let P_k be the Mtk associated with the Metropolis re-ordering algorithm targeting Φ_k. We prove the following lemma in Section D of Supplement I. The proof is constructed based on the fact that, in the Metropolis re-ordering algorithm, the next state depends on the current state (α,w,τ,u) only through α, which takes value in a finite set.

Lemma 12. For $k \in \mathsf{K}$, $\|P_k^2\|_{\Phi_k} < 1$.

With Lemma 12 in hand, it is now straightforward to establish the geometric ergodicity of the reversible jump algorithm.

Proposition 13. Suppose that $q_U(k) > 0$ for $k \in \mathsf{K}$, $q_B(k) > 0$ for $k < k_{\max}$, and $q_D(k) > 0$ for $k > 0$. Then the reversible jump chain is $L^2(\Pi)$ geometrically convergent and Π-a.e. geometrically ergodic.

Proof. Apply Theorem 1. Let P be the Mtk of the reversible jump chain. For $k \in \mathsf{K}$, (i) and (iii) in (H1) hold with $c_k = q_U(k)$. By Lemma 12, (ii) in (H1) holds with $t_0 = 2$.

To verify (H2), note that $\bar{P}(k,\{k'\}) > 0$ whenever $|k-k'| \le 1$. It is then evident that $\bar{P}$ is irreducible.

Thus, P is $L^2(\Pi)$ geometrically convergent. By Theorem 1 of Roberts and Tweedie (2001), it is Π-a.e. geometrically ergodic. □

Remark 14. The reversible jump algorithm can be further improved by adding more sophisticated between-model move types such as split and merge (Richardson and Green 1997; Zhang et al. 2004). Geometric convergence is preserved as long as the between-model movements remain irreducible.

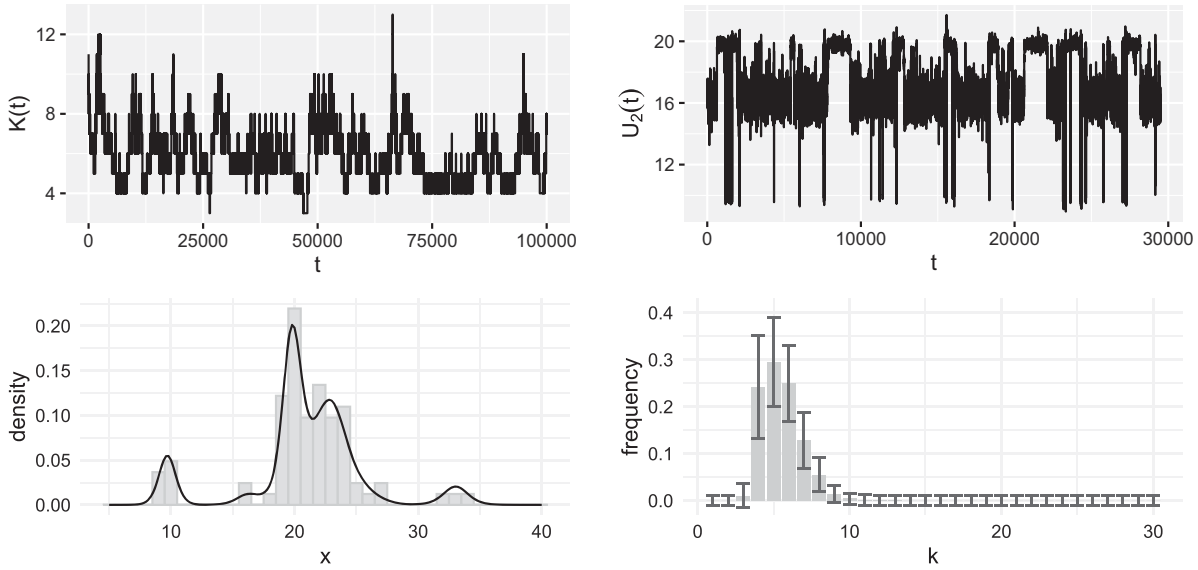

Figure 3. Top left: trace plot of $K(t)$; top right: trace plot of $U_2(t)$ when $K(t) = 5$; bottom left: histogram and predictive density for the galaxy dataset; bottom right: estimated posterior probabilities of $K = k$ and their 95% simultaneous confidence intervals. Hyperparameters: $k_{\max} = 30$, $\gamma = 2$, $b = 2$, $c = 3$, $\tau_0 = 1000$, $u_0 = 20$. $f_K(k) \propto 1/2^k$.

Geometric ergodicity allows us to construct 95% simultaneous Wald confidence intervals for the posterior probabilities of $K = k$ for $k = 1, \ldots, k_{\max}$. This is shown for the galaxy dataset in Figure 3. We have added $(\log m)\sqrt{1/b_m^2 + b_m/m}$ to the estimated asymptotic variances of the estimated probabilities, as in Section 4.2.4.

5. Discussion

$L^2(\Pi)$ geoemtrically convergence implies Π-a.e. geometric ergodicity. Harris recurrence, a regularity condition commonly used in MCMC analysis, can be enforced provided that we restrict our attention to some absorbing set (Meyn and Tweedie 2012, Theorem 9.0.1). See also Roberts and Rosenthal (2006) for conditions for Harris ergodicity in the context of trans-dimensional chains.

Under geometric ergodicity, one may use methods from for example, Jones et al. (2006) and Vats, Flegal, and Jones (2019) to construct confidence regions for uncertainty quantification. In our examples, we used Bonferroni correction to construct multiple Wald confidence intervals, but more sophisticated methods exist (Vats, Flegal, and Jones 2019; Robertson et al. 2020).

Some of the formulas in the proof of Lemma 3 (found in Section B of Supplement I) can be formulated in terms of Dirichlet forms. Dirichlet forms may be used to study Markov chains in terms of the conductance (Lawler and Sokal 1988), Peskun-Tierney ordering (Andrieu and Livingstone 2021), and functional inequalities (Power et al. 2024).

An obvious avenue for future research is obtaining practical quantitative convergence bounds for trans-dimensional chains. In particular, ways of computing or estimating $\mathrm{Gap}_{\bar{\Pi}}(\overline{P^* P})$ and $\mathrm{Gap}_{\bar{\Pi}}(\bar{P})$ would be useful for selecting good proposal distributions in a reversible jump algorithm. Whether our convergence bounds can be further sharpened and extended to the case where K is infinite is also of interest.

Supplementary Materials

Supplement I: Minor Results and Technical Proofs. This document contains a proof for the assertion that Π-irreducibility implies (H2), proofs for Lemmas 3, 11, 12, and simulation results concerning Monte Carlo confidence intervals.

Supplement II: Autoregression with Laplace errors. This document contains an application of the theory and methods herein to an autoregressive model with Laplace errors and unknown model order.

Acknowledgments

We thank the Editor, an anonymous Associate Editor, and three anonymous Referees for their helpful comments.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The author was partially supported by NSF DMS-2112887.

ORCID

Qian Qin http://orcid.org/0000-0002-6340-2977

References

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679. [6]

Andrieu, C., and Doucet, A. (1999), "Peskun-Tierney Ordering for Markovian Monte Carlo: Beyond the Reversible Scenario," *IEEE Transactions on Signal Processing*, 47, 2667–2676. [1]

Andrieu, C., and Livingstone, S. (2021), "Peskun-Tierney Ordering for Markov Chain and Process Monte Carlo: Beyond the Reversible Scenario," *Annals of Statistics*, 49, 1958–1981. [10]

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003), "Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions," *Journal of the Royal Statistical Society*, Series B, 65, 3–39. [7]

Caracciolo, S., Pelissetto, A., and Sokal, A. D. (1992), "Two Remarks on Simulated Tempering," Unpublished manuscript. [1,3]

Chakraborty, S., and Khare, K. (2017), "Convergence Properties of Gibbs Samplers for Bayesian Probit Regression with Proper Priors," *Electronic Journal of Statistics*, 11, 177–210. [7]

Chevallier, A., Fearnhead, P., and Sutton, M. (2023), "Reversible Jump PDMP Samplers for Variable Selection," *Journal of the American Statistical Association*, 118, 2915–2927. [1]

Chlebicka, I., Łatuszyński, K., and Miasojedow, B. (2024), "Solidarity of Gibbs Samplers: The Spectral Gap," *Annals of Applied Probability*, in press. [3]

Choi, M. C. (2020), "Metropolis–Hastings Reversiblizations of Non-Reversible Markov Chains," *Stochastic Processes and their Applications*, 130, 1041–1073. [4]

Cornish, N. J., and Littenberg, T. B. (2015), "Bayeswave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches," *Classical and Quantum Gravity*, 32, 135012. [1]

Ehlers, R., and Brooks, S. (2002), "Efficient Construction of Reversible Jump MCMC Proposals for Autoregressive Time Series Models," Technical Report. [1]

Fill, J. A. (1991), "Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process," *Annals of Applied Probability*, 1, 62–87. [4]

Flegal, J. M., and Jones, G. L. (2010), "Batch Means and Spectral Variance Estimators in Markov Chain Monte Carlo," *Annals of Statistics*, 38, 1034–1070. [7]

Gallegos-Herrada, M. A., Ledvinka, D., and Rosenthal, J. S. (2023), "Equivalences of Geometric Ergodicity of Markov Chains," *Journal of Theoretical Probability*, 37, 1230–1256. [1,3]

Ge, R., Lee, H., and Risteski, A. (2018), "Simulated Tempering Langevin Monte Carlo II: An Improved Proof Using Soft Markov Chain Decomposition," arXiv preprint. [2]

Geyer, C. J., and Møller, J. (1994), "Simulation Procedures and Likelihood Inference for Spatial Point Processes," *Scandinavian Journal of Statistics*, 21, 359–373. [1]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [1,9]

Guan, Y., and Krone, S. M. (2007), "Small-World MCMC and Convergence to Multi-Modal Distributions: From Slow Mixing to Fast Mixing," *Annals of Applied Probability*, 17, 284–304. [3]

Hairer, M. (2006), "Ergodic Properties of Markov Processes," Lecture notes, Univ. Warwick. [5]

Hopkins, M., Reeber, E., Forman, G., and Suermondt, J. (1999), "Spambase," UCI Machine Learning Repository. DOI:10.24432/C53G6X. [7]

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006), "Fixed-Width Output Analysis for Markov Chain Monte Carlo," *Journal of the American Statistical Association*, 101, 1537–1547. [7,10]

Jones, G. L., and Hobert, J. P. (2001), "Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo," *Statistical Science*, 16, 312–334. [1]

Lawler, G. F., and Sokal, A. D. (1988), "Bounds on the l^2 'Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger's Inequality," *Transactions of the American Mathematical Society*, 309, 557–580. [10]

Madras, N., and Randall, D. (2002), "Markov Chain Decomposition for Convergence Rate Analysis," *Annals of Applied Probability*, 12, 581–606. [1,2,3,4]

Meyn, S. P., and Tweedie, R. L. (2012), *Markov Chains and Stochastic Stability* (2nd ed.), London: Springer. [10]

Ortner, M., Descombes, X., and Zerubia, J. (2006), "A Reversible Jump MCMC Sampler for Object Detection in Image Processing," in *Monte Carlo and Quasi-Monte Carlo Methods 2004*, eds. A. Keller, pp. 389–401, Cham: Springer. [1]

Paulin, D. (2015), "Concentration Inequalities for Markov Chains by Marton Couplings and Spectral Methods," *Electronic Journal of Probability*, 20, 1–32. [4]

Power, S., Rudolf, D., Sprungk, B., and Wang, A. Q. (2024), "Weak Poincaré Inequality Comparisons for Ideal and Hybrid Slice Sampling," arXiv preprint. [10]

Qin, Q., and Jones, G. L. (2022), "Convergence Rates of Two-Component MCMC Samplers," *Bernoulli*, 28, 859–885. [3]

Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components," (with Discussion), *Journal of the Royal Statistical Society*, Series B, 59, 731–792. [8,9]

Roberts, G. O., and Rosenthal, J. S. (1997), "Geometric Ergodicity and Hybrid Markov Chains," *Electronic Communications in Probability*, 2, 13–25. [1,3]

——— (2006), "Harris Recurrence of Metropolis-within-Gibbs and Trans-Dimensional Markov Chains," *Annals of Applied Probability*, 16, 2123–2139. [1,10]

Roberts, G. O., and Tweedie, R. L. (2001), "Geometric L^2 and L^1 Convergence are Equivalent for Reversible Markov Chains," *Journal of Applied Probability*, 38, 37–41. [1,3,7,9]

Robertson, N., Flegal, J. M., Vats, D., and Jones, G. L. (2020), "Assessing and Visualizing Simultaneous Simulation Error," *Journal of Computational and Graphical Statistics*, 30, 324–334. [10]

Roeder, K. (1990), "Density Estimation with Confidence Sets Exemplified by Superclusters and Voids in the Galaxies," *Journal of the American Statistical Association*, 85, 617–624. [9]

Roy, V., and Hobert, J. P. (2007), "Convergence Rates and Asymptotic Standard Errors for Markov Chain Monte Carlo Algorithms for Bayesian Probit Regression," *Journal of the Royal Statistical Society*, Series B, 69, 607–623. [7]

Schreck, A., Fort, G., Le Corff, S., and Moulines, E. (2015), "A Shrinkage-Thresholding Metropolis Adjusted Langevin Algorithm for Bayesian Variable Selection," *IEEE Journal of Selected Topics in Signal Processing*, 10, 366–375. [1]

Troughton, P. T., and Godsill, S. J. (1998), "A Reversible Jump Sampler for Autoregressive Time Series," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)* (Vol. 4), pp. 2257–2260. IEEE. [1]

Vats, D., Flegal, J. M., and Jones, G. L. (2019), "Multivariate Output Analysis for Markov Chain Monte Carlo," *Biometrika*, 106, 321–337. [1,10]

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. (2004), "Reversible Jump Markov Chain Monte Carlo Strategies for Bayesian Model Selection in Autoregressive Processes," *Journal of Time Series Analysis*, 25, 785–809. [1]

Woodard, D. B., Schmidler, S. C., and Huber, M. (2009), "Conditions for Rapid Mixing of Parallel and Simulated Tempering on Multimodal Distributions," *Annals of Applied Probability*, 19, 617–640. [2,3]

Zhang, Z., Chan, K. L., Wu, Y., and Chen, C. (2004), "Learning a Multivariate Gaussian Mixture Model with the Reversible Jump MCMC Algorithm," *Statistics and Computing*, 14, 343–355. [9]