Analysis of the Pólya-Gamma block Gibbs sampler for Bayesian logistic linear mixed models

Xin Wang∗ and Vivekananda Roy†
Department of Statistics, Iowa State University, Ames, IA

November, 2017

Abstract
In this article, we construct a two-block Gibbs sampler using Polson et al.’s (2013) data augmentation technique with Pólya-Gamma latent variables for Bayesian logistic linear mixed models under proper priors. Furthermore, we prove the uniform ergodicity of this Gibbs sampler, which guarantees the existence of the central limit theorems for MCMC based estimators.

key words: Data augmentation, Markov Chain, Logit link, Pólya-Gamma distribution, Uniform ergodicity

1 Introduction

Consider the logistic linear mixed model set-up (Charles E. McCulloch and Neuhaus, 2008; McCulloch, 2003). Let \((Y_1, Y_2, \ldots, Y_N)\) denote the vector of Binomial\((n_i, p_i)\) random variables, \(x_i\) and \(z_i\) be the \(p \times 1\) and \(q \times 1\) known covariates and random effect design vectors respectively associated with the \(i\)th observation for \(i = 1, \ldots, N\). Let \(\beta \in \mathbb{R}^p\) be the unknown vector of regression coefficients and \(u \in \mathbb{R}^q\) be the random effects vector. Assume that \(p_i = F(x_i^T \beta + z_i^T u)\), where \(F\) is the standard logistic distribution function, that is \(F(t) \equiv e^t/(1 + e^t)\) for \(t \in \mathbb{R}\). Suppose we have \(r\) random effects with \(u = (u_1^T, \ldots, u_r^T)^T\), where \(u_j\) is a \(q_j \times 1\) vector with \(q_1 + \cdots + q_r = q\), and \(u_j \overset{\text{ind}}{\sim} N(0, I_{q_j}/\tau_j)\), where \(\tau_j \in \mathbb{R}_+ = (0, \infty)\) is the precision parameter associated with \(u_j\) for \(j = 1, \ldots, r\). The joint distribution of \(u\) is \(N(0, D(\tau)^{-1})\), where \(D(\tau) = \oplus_{j=1}^r \tau_j I_{q_j}\), \(\tau = (\tau_1, \ldots, \tau_r)\) and \(\oplus\) denotes the direct sum. The data model for the logistic linear mixed model is

\[
\begin{align*}
Y_i|\beta, u & \overset{\text{ind}}{\sim} \text{Binomial}(n_i, p_i) \text{ for } i = 1, \ldots, N \text{ with } \\
p_i & = F(x_i^T \beta + z_i^T u) \text{ for } i = 1, \ldots, N, \\
u_j|\tau_j & \overset{\text{ind}}{\sim} N\left(0, \frac{1}{\tau_j} I_{q_j}\right), j = 1, \ldots, r.
\end{align*}
\]

∗Email: xinwang@iastate.edu
†Email: vroy@iastate.edu
Let \( \mathbf{y} = (y_1, y_2, \ldots, y_N)^T \) be the observed Binomial response variables. The likelihood function for \((\mathbf{\beta}, \mathbf{\tau})\) is
\[
L(\mathbf{\beta}, \mathbf{\tau}|\mathbf{y}) = \int_{\mathbb{R}^q} \prod_{i=1}^N \binom{n_i}{y_i} \left[ \frac{\exp \left( x_i^T \mathbf{\beta} + z_i^T \mathbf{u} \right)}{1 + \exp \left( x_i^T \mathbf{\beta} + z_i^T \mathbf{u} \right)} \right]^{y_i} \phi_q \left( \mathbf{u}; 0, D(\mathbf{\tau})^{-1} \right) d\mathbf{u},
\]
where \( \phi_q(s; a, B) \) is the probability density function of the \( q \)-dimensional normal distribution with mean vector \( a \) and covariance matrix \( B \) evaluated at \( s \). In Bayesian framework, let \( \pi(\mathbf{\beta}) \) and \( \pi(\mathbf{\tau}) \) be the prior densities for \( \mathbf{\beta} \) and \( \mathbf{\tau} \) respectively. Assume that \( \mathbf{\beta} \) and \( \mathbf{\tau} \) are apriori independent. The joint posterior density of \((\mathbf{\beta}, \mathbf{\tau})\) is
\[
\pi(\mathbf{\beta}, \mathbf{\tau}|\mathbf{y}) = \frac{1}{c_0(\mathbf{y})} L(\mathbf{\beta}, \mathbf{\tau}|\mathbf{y}) \pi(\mathbf{\beta}) \pi(\mathbf{\tau}),
\]
where \( c_0(\mathbf{y}) \) is the marginal density of \( \mathbf{y} \) with \( c_0(\mathbf{y}) = \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} L(\mathbf{\beta}, \mathbf{\tau}|\mathbf{y}) \pi(\mathbf{\beta}) \pi(\mathbf{\tau}) d\mathbf{\beta} d\mathbf{\tau} \).

The posterior density (3) is intractable for any choice of the prior distributions of \( \mathbf{\beta} \) and \( \mathbf{\tau} \). Generally, Markov chain Monte Carlo (MCMC) algorithms are used for exploring these posterior densities. Even in the absence of random effects, for generalized linear models, MCMC algorithms are needed to summarize the associated posterior densities. For probit regression models, Albert and Chib (1993) proposed a widely used data augmentation (DA) algorithm to sample from the corresponding posterior distributions. Roy and Hobert (2007) and Chakraborty and Khare (2017) proved the geometric ergodicity of this DA algorithm for Bayesian probit regression model under improper and proper priors respectively. Wang and Roy (2017) recently extended the convergence rate analysis of the block Gibbs samplers based on this DA technique for Bayesian probit linear mixed models under both proper and improper priors.

For logistic regression models, there have been several attempts for producing a DA algorithm similar to Albert and Chib’s (1993) algorithm for the probit regression model (see e.g. Holmes and Held (2006) and Frühwirth-Schnatter and Frühwirth (2010)). Unfortunately, these algorithms are far more complex than Albert and Chib’s (1993) algorithm. Only recently, Polson et al. (2013) produced such a DA algorithm for logistic regression models using Pólya-Gamma latent variables. Choi and Hobert (2013) proved uniform ergodicity of the Pólya-Gamma DA Markov chain under normal priors on the regression parameters. Choi and Román (2017) showed that the Markov operator based on Polson et al.’s (2013) DA algorithm for one-way logistic ANOVA model is trace-class, which implies that the associated Markov Chain is geometrically ergodic. Both Choi and Hobert (2013) and Choi and Román (2017) considered the special case when the data are binary, that is \( n_i = 1 \) for all \( i \). However, there is no result in the literature about convergence analysis of any Gibbs samplers for Bayesian logistic linear mixed models. In this article, we construct a two-block Gibbs sampler for Bayesian logistic linear mixed models with normal priors on regression parameters and truncated Gamma priors on precision parameters. We further establish uniform ergodicity of this Gibbs sampler.

The article is organized as follows. In section 2, we construct the two-block Gibbs sampler for the Bayesian logistic linear mixed model under proper priors. In section 3,
we prove the uniform ergodicity of the underlying Markov chain. Finally, we have some discussions in section 4.

2 Two-block Gibbs sampler

In Polson et al. (2013), a logistic linear mixed model example is introduced. In their example, normal distribution is used as the prior for regression coefficients and Gamma distribution is used as the prior for precision parameters. We assume the following priors: \( \beta \sim N_p(Q^{-1}\mu_0, Q^{-1}) \) for some \( p \times p \) positive definite matrix \( Q \) and \( \mu_0 \in \mathbb{R}^p \), \( \tau_j \sim \text{truncated Gamma}(a_j, b_j, \tau_0) \), \( j = 1, \ldots, r \), where \( b_j > 0 \) for \( j = 1, \ldots, r \). The density function of truncated Gamma \( \text{Gamma}(a_j, b_j, \tau_0) \) is

\[
f(\tau_j|a_j, b_j, \tau_0) = |c(\tau_0, a_j, b_j)|^{-1}\tau_j^{a_j-1} \exp(-b_j\tau_j) I(\tau_j \geq \tau_0),
\]

where \( c(\tau_0, a_j, b_j) = \int_0^\infty \tau_j^{a_j-1} \exp(-b_j\tau) \, d\tau \) and \( \tau_0 > 0 \) is a known constant.

By Theorem 1 in Polson et al. (2013),

\[
\frac{\left[ \exp (x_i^T \beta + z_i^T u) \right]^{y_i}}{[1 + \exp (x_i^T \beta + z_i^T u)]^{\tau_i}} = 2^{-n_i} \exp \left[ \kappa_i \left( x_i^T \beta + z_i^T u \right) \right] \times \int_0^\infty \exp \left[ -\omega_i \left( x_i^T \beta + z_i^T u \right)^2 / 2 \right] p(\omega_i) \, d\omega_i,
\]

where \( p(\omega_i) \) is the probability density function of the random variable \( \omega_i \sim \text{PG}(n_i, 0) \) and \( \kappa_i = y_i - n_i/2 \) for \( i = 1, \ldots, N \). Here, PG\((n_i, 0)\) denotes the Polya-Gamma distribution with parameters \( n_i \) and 0 with density

\[
f(x|n_i, 0) = \frac{2^{n_i-1}}{\Gamma(n_i)} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(n + n_i)(2n + n_i)}{\Gamma(n + 1)} e^{- \frac{(2n + n_i)^2}{4x}} \quad x > 0.
\]

Let \( \omega = (\omega_1, \ldots, \omega_N) \) and the joint (posterior) density of \( \beta, u, \omega \) and \( \tau \) be

\[
\pi(\beta, u, \omega, \tau | y) \propto \prod_{i=1}^N \exp \left[ \kappa_i \left( x_i^T \beta + z_i^T u \right) - \omega_i \left( x_i^T \beta + z_i^T u \right)^2 / 2 \right] p(\omega_i)
\]

\[
\times \phi_q \left( u; 0, D(\tau)^{-1} \right) \phi_p \left( \beta; Q^{-1}\mu_0, Q^{-1} \right) \prod_{j=1}^r \tau_j^{a_j-1} e^{-b_j \tau_j} I(\tau_j \geq \tau_0).
\]

From (2), (3) and (5), it follows that \( \int_{\mathbb{R}^n} \int_{\mathbb{R}^N} \pi(\beta, u, \omega, \tau | y)d\omega du = \pi(\beta, \tau | y) \), which is our target posterior density. Using draws from all full conditional distribution distributions of (6), we can run a Gibbs sampler with stationary density (6). It is known that by combining and simultaneously drawing multiple parameters, the convergence of the Gibbs sampler can be improved (Liu et al., 1994), although the “blocking” to be computationally efficient, the corresponding joint conditional distributions need to be tractable. Here we construct a two-block Gibbs sample for (6). Let \( \eta = (\beta^T, u^T)^T \),
\[ \kappa = (\kappa_1, \ldots, \kappa_n) \text{ and } M = (X, Z) \text{ with } \text{ith row } m_i^T \text{ and } \Omega \text{ be the } n \times n \text{ diagonal matrix with ith diagonal element } \omega_i. \] Standard calculations show that the conditional density of \( \eta \) is

\[
\pi(\eta | \omega, \tau, y) \propto \exp \left[ -\frac{1}{2} \eta^T \kappa^T M \Omega M \eta + \eta^T M^T \kappa \right] \exp \left[ -\frac{1}{2} \eta^T A(\tau) \eta + \eta^T l \right],
\]

where \( l = (\mu_0^T, 0_{1 \times q})^T \), and \( A(\tau) = Q \oplus D(\tau) \). That is,

\[
\eta | \omega, \tau, y \sim N \left( \Sigma^{-1} \mu, \Sigma^{-1} \right) \tag{7}
\]

Similarly, the conditional density of \( (\omega, \tau) \) is

\[
\pi(\omega, \tau | \beta, u, y) \propto \prod_{i=1}^N \exp \left[ -\omega_i \left( x_i^T \beta + z_i^T u \right)^2 / 2 \right] p(\omega_i)
\]
\[
\times \prod_{j=1}^r \tau_j^{a_j + q_j / 2 - 1} \exp \left[ -\left( u_j^T u_j / 2 + b_j \right) \tau_j \right] I(\tau_j \geq \tau_0).
\]

So, given \( \beta, u, y \), we have that \( \omega \) and \( \tau \) are conditionally independent with

\[
\omega_i | \eta, y \overset{\text{iid}}{\sim} PG \left( n_i, |m_i^T \eta| \right), \quad i = 1, \ldots, N,
\]
\[
\tau_j | \eta, y \overset{\text{iid}}{\sim} \text{truncated Gamma} \left( a_j + \frac{q_j}{2}, b_j + \frac{u_j^T u_j}{2}, \tau_0 \right), \quad j = 1, \ldots, r.
\]

**Remark 1.** As in *Wang and Roy (2017)*, we assume that the prior distribution for \( \tau_j \) is a truncated Gamma distribution. However, while implementing the block Gibbs sampler in practice, a number slightly larger than the machine precision zero can be treated as \( \tau_0 \), practically avoiding the need to use any rejection sampling algorithms to draw from the truncated conditional distribution of \( \tau \).

Thus, one single iteration of the block Gibbs sampler \( \{ \eta^{(m)}, \omega^{(m)}, \tau^{(m)} \}_{m=0}^\infty \) has the following two steps:

**Algorithm:** The \((m + 1)\)st iteration for the two-block Gibbs sampler

1. Draw \( \tau_j^{(m+1)} \) from truncated Gamma \( \left( a_j + \frac{q_j}{2}, b_j + \frac{u_j^T u_j}{2}, \tau_0 \right) \) with \( u = u^{(m)} \) for \( j = 1, \ldots, r \), and independently draw \( \omega_i^{(m+1)} \overset{\text{iid}}{\sim} PG \left( n_i, |m_i^T \eta^{(m)}| \right) \) for \( i = 1, \ldots, N \).

2. Draw \( \eta^{(m+1)} \) from \( \eta^{(m+1)} \sim N_{p+q} \left( \Sigma^{(m)-1} \left( M^T \kappa + l \right), \Sigma^{(m)-1} \right) \), where \( \Sigma^{(m)} = M^T \Omega^{(m+1)} M + A\left( \tau^{(m+1)} \right) \) and the diagonal elements of \( \Omega^{(m+1)} \) are \( \omega_i^{(m+1)} \), \( i = 1, \ldots, N \).

*Polson et al. (2013)* developed an efficient method for sampling from PG distribution, which is the only nonstandard distribution involved in the above Gibbs sampler.
3 Uniform ergodicity of the two-block Gibbs sampler

In this section, we prove the uniform ergodicity of the two-block Gibbs sampler \( \{ \eta^{(m)}, \omega^{(m)}, \tau^{(m)} \}_{m=0}^{\infty} \), which has the same rate of convergence as the \( \eta \)-marginal Markov chain \( \{ \eta^{(m)} \}_{m=0}^{\infty} \) (Roberts and Rosenthal, 2001). Below we analyze the \( \Psi \equiv \{ \eta^{(m)} \}_{m=0}^{\infty} \) chain.

Let \( \eta' \) be the current state and \( \eta \) be the next state, then the Markov transition density (Mtd) of \( \Psi \) is

\[
k(\eta|\eta') = \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} \pi(\eta, \omega, \tau, y) \pi(\omega, \tau|\eta', y) d\omega d\tau,
\]

where \( \pi(\cdot|\cdot, y) \)'s are the conditional densities from section 2. Routine calculations show that \( k(\eta|\eta') \) is reversible and thus invariant with respect to the marginal density of \( \eta \) denoted as \( \pi(\eta|y) \equiv \int_{\mathbb{R}^q} \pi(\eta, \omega, \tau|y) d\omega d\tau \), where \( \pi(\eta, \omega, \tau|y) \) is defined in (6). Since \( k(\eta|\eta') \) is strictly positive, the Markov chain \( \Psi \) is Harris ergodic (Hobert, 2011).

Let \( \mathcal{B} \) denote the Borel \( \sigma \)-algebra of \( \mathbb{R}^{p+q} \) and \( K(\cdot, \cdot) \) be the Markov transition function corresponding to the Mtd \( k(\cdot, \cdot) \) in (8), that is, for any set \( A \in \mathcal{B} \), \( \eta', \eta'' \in \mathbb{R}^{p+q} \) and any \( j = 0, 1, \ldots, \)

\[
K(\eta', A) = \Pr(\eta'^{(j+1)} \in A|\eta'^{(j)} = \eta') = \int_A k(\eta|\eta') d\eta.
\]

Then the \( m \)-step Markov transition function is \( K^m(\eta', A) = \Pr(\eta'^{(m+j)} \in A|\eta'^{(j)} = \eta') \). Let \( \Pi(\cdot|y) \) be the probability measure with density \( \pi(\eta|y) \). The Markov chain \( \Psi \) is geometrically ergodic if there exists a constant \( 0 < \theta < 1 \) and a function \( G : \mathbb{R}^{p+q} \to \mathbb{R}^+ \) such that for any \( \eta \in \mathbb{R}^{p+q}, \)

\[
||K^m(\eta, \cdot) - \Pi(\cdot|y)|| := \sup_{A \in \mathcal{B}} |K^m(\eta, A) - \Pi(A|y)| \leq G(\eta)t^m.
\]

If \( G(\eta) \) is bounded above, then the corresponding Markov chain is uniformly ergodic. The following theorem establishes uniform ergodicity of the Markov chain \( \Psi \) by constructing a minorization condition.

**Theorem 1.** Assume that \( a_j + q_j/2 \geq 1 \) and \( b_j > 0 \) for all \( j \), then the Markov chain \( \Psi \) is uniformly ergodic.

**Proof.** We show that there exists a \( \delta > 0 \) and a density function \( h : \mathbb{R}^{p+q} \to [0, \infty) \) such that, for all \( \eta', \eta \in \mathbb{R}^{p+q}, \)

\[
k(\eta|\eta') \geq \delta h(\eta).
\]

By Roberts and Rosenthal (2004)[Theorem 8], (11) implies that the Markov chain \( \Psi \) is uniformly ergodic. Furthermore, under (11), (10) holds with \( G = 1 \) and \( t = 1 - \delta \).

For \( \tau_j \geq \tau_0, j = 1, \ldots, q, \) \( \Sigma \geq \bar{M}^T\Omega M + A(\tau_0) \), that is \( \Sigma - (\bar{M}^T\Omega M + A(\tau_0)) \) is positive semidefinite. So \( |\Sigma| \geq |\bar{M}^T\Omega M + A(\tau_0)| = |A(\tau_0)| |\bar{M}^T\Omega M + I| \geq
\[ |\mathbf{A}(\tau_0)|, \text{ where } \mathbf{M} = \mathbf{M}(\tau_0)^{-1/2}. \text{ And } \mu^T \Sigma^{-1} \mu \leq \mu^T (\mathbf{M}^T \Sigma \mathbf{M} + \mathbf{A}(\tau_0))^{-1} \mu \leq \mu^T \mathbf{A}(\tau_0)^{-1} \mu. \text{ Then,}
\]
\[
\pi(\eta|\omega, \tau, y) = (2\pi)^{-\frac{p+n}{2}} |\Sigma|^{\frac{p}{2}} \exp \left[ -\frac{1}{2} \left( \eta - \Sigma^{-1} \mu \right)^T \Sigma \left( \eta - \Sigma^{-1} \mu \right) \right] 
\]
\[
\geq (2\pi)^{-\frac{p+n}{2}} |\mathbf{A}(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \left( \eta^T \Sigma \eta - 2\eta^T \mu \right) \right] 
\]
\[
= (2\pi)^{-\frac{p+n}{2}} |\mathbf{A}(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \beta^T Q \beta - \frac{1}{2} \eta^T D(\tau) \eta + \eta^T \mu - \frac{1}{2} \mu^T \mathbf{A}(\tau_0)^{-1} \mu \right] 
\]
\[
\times \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \omega_i \left( m_i^T \eta \right)^2 \right].
\]

Therefore,
\[
\pi(\eta|\omega, \tau, y) \pi(\omega, \tau', y)
\]
\[
\geq (2\pi)^{-\frac{p+n}{2}} |\mathbf{A}(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \beta^T Q \beta - \frac{1}{2} \eta^T D(\tau) \eta + \eta^T \mu - \frac{1}{2} \mu^T \mathbf{A}(\tau_0)^{-1} \mu \right] 
\]
\[
\times \prod_{j=1}^{p} \frac{1}{c \left( \tau_0, a_j + q_j/2, b_j + u_j^T u_j'/2 \right)} \tau_j^{q_j/2 - 1} \exp \left[ - \left( b_j + u_j^T u_j'/2 \right) \tau_j \right] I(\tau_j \geq \tau_0) 
\]
\[
\times \prod_{i=1}^{N} \cosh n_i \left( \frac{m_i^T \eta'}{2} \right) \exp \left[ - \left( \frac{m_i^T \eta'}{2} \right)^2 + \left( \frac{m_i^T \eta}{2} \right)^2 \omega_i \right] p(\omega_i). 
\]

According to Polson et al. (2013) and Choi and Hobert (2013),
\[
\int_{\mathbb{R}_+} \exp \left[ - \left( \frac{m_i^T \eta'}{2} \right)^2 + \left( \frac{m_i^T \eta}{2} \right)^2 \omega_i \right] p(\omega_i) d\omega_i = \left[ \cosh \left( \frac{\sqrt{\left( m_i^T \eta' \right)^2 + \left( m_i^T \eta \right)^2}}{2} \right) \right]^{-n_i}
\]
\[
\geq \left[ \cosh \left( \frac{m_i^T \eta'}{2} + \frac{m_i^T \eta}{2} \right) \right]^{-n_i} \geq \left[ 2 \cosh \left( \frac{m_i^T \eta'}{2} \right) \cosh \left( \frac{m_i^T \eta}{2} \right) \right]^{-n_i},
\]

implying
\[
\cosh n_i \left( \frac{m_i^T \eta'}{2} \right) \int_{\mathbb{R}_+} \exp \left[ - \left( \frac{m_i^T \eta'}{2} \right)^2 + \left( \frac{m_i^T \eta}{2} \right)^2 \omega_i \right] p(\omega_i) d\omega_i \geq 2^{-n_i} \cosh^{-n_i} \left( \frac{m_i^T \eta}{2} \right)
\]
\[
\geq 2^{-n_i} \left[ \exp \left( \frac{m_i^T \eta}{2} \right) \right]^{-n_i} \geq \exp \left( -\frac{n_i}{4} \left( m_i^T \eta \right)^2 \right) = 2^{-n_i} e^{-n_i/4} \exp \left[ -\frac{n_i}{4} \left( m_i^T \eta \right)^2 \right].
\]
So we have,

\[
\int_{\mathbb{R}^d_+} \pi(\eta|\omega, \tau, y) \pi(\omega, \tau|\eta', y) d\omega \\
\geq (2\pi)^{-\frac{d+2}{2}} |A(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \beta^T Q \beta + \eta^T \mu - \frac{1}{2} \mu^T A(\tau_0)^{-1} \mu \right] \\
\times 2^{-n} e^{-\frac{2}{4}} \exp \left[ -\frac{1}{4} \eta^T M^T \Lambda M \eta \right] \\
\times \prod_{j=1}^{r} \left[ c \left( \tau_0, a_j + q_j/2, b_j + u_j^T u_j'/2 \right) \right]^{1/2} \exp \left[ -\left( b_j + u_j^T u_j'/2 + u_j^T u_j/2 \right) \tau_j \right] I(\tau_j \geq \tau_0),
\]

where \( n = \sum_{i=1}^{N} n_i \) and \( \Lambda \) is the \( N \times N \) diagonal matrix with \( i \)-th diagonal element \( n_i \).

Thus,

\[
k(\eta|\eta') = \int_{\mathbb{R}^d_+} \int_{\mathbb{R}^d_+} \pi(\eta|\omega, \tau, y) \pi(\omega, \tau|\eta', y) d\omega d\tau \\
\geq (2\pi)^{-\frac{d+4}{2}} |A(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \beta^T Q \beta + \eta^T \mu - \frac{1}{2} \mu^T A(\tau_0)^{-1} \mu \right] \\
\times 2^{-n} e^{-\frac{2}{4}} \exp \left[ -\frac{1}{4} \eta^T M^T \Lambda M \eta \right] \\
\times \prod_{j=1}^{r} \left[ c \left( \tau_0, a_j + q_j/2, b_j + u_j^T u_j'/2 \right) \right]^{1/2} \exp \left[ -\left( b_j + u_j^T u_j'/2 + u_j^T u_j/2 \right) \tau_j \right] d\tau_j.
\]

Now consider

\[
\frac{1}{c \left( \tau_0, a_j + q_j/2, b_j + u_j^T u_j'/2 \right)} \int_{\tau_0}^{\infty} \tau_j^{a_j+q_j/2-1} \exp \left[ -\left( b_j + u_j^T u_j'/2 + u_j^T u_j/2 \right) \tau_j \right] d\tau_j
\]

\[
= \frac{\left( b_j + u_j^T u_j'/2 \right)^{a_j+q_j/2}}{\left( b_j + u_j^T u_j'/2 + u_j^T u_j/2 \right)^{a_j+q_j/2}} \frac{\int_{\tau_0}^{\infty} \tau_j^{a_j+q_j/2-1} \exp (-x) dx}{\int_{b_j + u_j^T u_j'/2}^{\infty} \tau_j^{a_j+q_j/2-1} \exp (-x) dx}.
\]

For \( x \geq 0 \), define \( f_1(x) = \int_{b_j + u_j^T u_j'/2}^{\infty} \tau_j^{a_j+q_j/2-1} \exp (-t) dt \) and \( f_2(x) = \int_{b_j + x}^{\infty} \tau_j^{a_j+q_j/2-1} \exp (-t) dt \) and \( g(x) = \int_{b_j + x}^{\infty} \tau_j^{a_j+q_j/2-1} \exp (-t) dt \) and \( g(x) = \lim_{x \to \infty} g(x) = 0 \). Thus \( f_1(x) / f_2(x) \geq \exp \left[ -\tau_0 u_j^T u_j/2 \right] \). Also \( (b_j + u_j^T u_j'/2)/(b_j + u_j^T u_j'/2 + u_j^T u_j/2) \geq b_j/(b_j + u_j^T u_j/2) \).

So

\[
\kappa(\eta|\eta') \geq (2\pi)^{-\frac{d+4}{2}} |A(\tau_0)|^{1/2} \exp \left[ -\frac{1}{2} \beta^T Q \beta + \eta^T \mu - \frac{1}{2} \mu^T A(\tau_0)^{-1} \mu \right] \\
\times 2^{-n} e^{-\frac{2}{4}} \exp \left[ -\frac{1}{4} \eta^T M^T \Lambda M \eta \right] \prod_{j=1}^{r} \left( \frac{b_j}{b_j + u_j^T u_j/2} \right)^{a_j+q_j/2} \exp \left( -\tau_0 u_j^T u_j/2 \right).
\]
Let
\[ c_1(M, y) = \int_{\mathbb{R}^{p+q}} \exp \left[ -\frac{1}{2} \beta^T Q \beta + \eta^T \mu - \frac{1}{4} \eta^T M^T \Lambda M \eta \right] \times \prod_{j=1}^r \left( \frac{b_j}{b_j + u_j^T u_j/2} \right)^{a_j + q_j/2} \exp \left( -\tau_0 u_j^T u_j/2 \right) d\eta \leq (2\pi)^{-\frac{p+q}{2}} |A(\tau_0)|^{-1/2} \exp \left( \frac{1}{2} \mu^T A(\tau_0) \mu \right) 2^n e^{-\frac{n}{4}} < \infty. \]

So there exists a density function \( h(\eta) \) and \( \delta > 0 \) such that,
\[ k(\eta|\eta') = \int_{\mathbb{R}_+^r} \int_{\mathbb{R}_+^2} \pi(\eta|\omega, \tau, y) \pi(\omega, \tau|\eta', y) d\omega d\tau \geq \delta h(\eta), \]
where
\[ h(\eta) = \frac{1}{c_1(M, y)} \exp \left[ -\frac{1}{2} \beta^T Q \beta + \eta^T \mu - \frac{1}{4} \eta^T M^T \Lambda M \eta \right] \times \prod_{j=1}^r \left( \frac{b_j}{b_j + u_j^T u_j/2} \right)^{a_j + q_j/2} \exp \left( -\tau_0 u_j^T u_j/2 \right), \]
and \( \delta = (2\pi)^{-\frac{p+q}{2}} |A(\tau_0)|^{-1/2} 2^{-n} e^{-\frac{n}{4}} \cdot \exp \left[ -\frac{1}{2} \mu^T A(\tau_0)^{-1} \mu \right] c_1(M, y). \]

Hence the Markov chain is uniformly ergodic.

\[ \square \]

**Remark 2.** Since Theorem 1 does not put any conditions on \( y, X, Z, N, p \) and \( q \), it is applicable in high dimensional situations where \( p \) (or \( q \)) can be much larger than \( N \).

**Remark 3.** Following the proof of Theorem 1, the uniform ergodicity result in Choi and Hobert (2013) can be extended to binomial data.

**Remark 4.** Since \( 1/(x+1) \geq \exp(-x) \), we have \( b_j/(b_j + u_j^T u_j/2) \geq \exp \left[ -u_j^T u_j/(2b_j) \right] \).

Using this inequality, we have
\[ \delta \geq 2^{-n} e^{-\frac{n}{2}} |A(\tau_0)|^{1/2} |\Sigma_1|^{-1/2} \exp \left[ -\frac{1}{2} \mu^T A(\tau_0)^{-1} \mu + \frac{1}{2} \mu^T \Sigma_1^{-1} \mu \right], \]
where \( \Sigma_1 = \frac{1}{2} M^T \Lambda M + Q \oplus \left( \oplus_{j=1}^r \{\tau_0 + (a_j + q_j)/(2b_j)\} I_{q_j} \right) \). This in turn, gives a computable upper bound to the total variation distance to the stationary in (10).

### 4 Discussion

We prove uniform ergodicity of the two-block Gibbs sampler for Bayesian logistic linear mixed models, which guarantees the existence of central limit theorem for MCMC
estimators under a finite second moment condition (Jones, 2004). Thus, our result has important practical implications as it allows for obtaining valid asymptotic standard errors for the posterior estimates (Flegal and Jones, 2010). Convergence rates analysis of Gibbs samplers for Bayesian logistic linear mixed models with improper priors is a potential future project.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

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.

Charles E. McCulloch, S. R. S. and Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. Wiley, 2nd ed edition.

Choi, H. M. and Hobert, J. P. (2013). The Pólya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics*, 7:2054–2064.

Choi, H. M. and Román, J. C. (2017). Analysis of Pólya-Gamma Gibbs sampler for Bayesian logistic analysis of variance. *Electronic Journal of Statistics*, 11(1):326–337.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):1034–1070.

Frühwirth-Schnatter, S. and Frühwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In *Statistical Modelling and Regression Structures*, pages 111–132. Springer.

Hobert, J. P. (2011). The data augmentation algorithm: Theory and methodology. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1):145–168.

Jones, G. L. (2004). On the markov chain central limit theorem. *Probability surveys*, 1:299â€ť320.

Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, pages 27–40.

McCulloch, C. E. (2003). Generalized linear mixed models. *NSF-CBMS regional conference series in probability and statistics*, 7:i–84.
Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.

Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, 28:489–504.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71.

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(4):607–623.

Wang, X. and Roy, V. (2017). Convergence analysis of block Gibbs samplers for Bayesian probit linear mixed models. *arXiv preprint arXiv:1706.01846*.