Electronic Journal of Statistics
Vol. 11 (2017) 326–337
ISSN: 1935-7524
DOI: 10.1214/17-EJS1227

Analysis of Polya-Gamma Gibbs sampler for Bayesian logistic analysis of variance

Hee Min Choi

Department of Statistics
University of California, Davis
e-mail: hmchoi@ucdavis.edu

and

Jorge Carlos Román

Department of Mathematics and Statistics
San Diego State University
e-mail: jcroman@mail.sdsu.edu

Abstract: We consider the intractable posterior density that results when the one-way logistic analysis of variance model is combined with a flat prior. We analyze Polson, Scott and Windle's (2013) data augmentation (DA) algorithm for exploring the posterior. The Markov operator associated with the DA algorithm is shown to be trace-class.

AMS 2000 subject classifications: Primary 60J27; secondary 60K35.
Keywords and phrases: Polya-Gamma distribution, data augmentation algorithm, geometric convergence rate, Markov chain, Markov operator, Monte Carlo, trace-class operator.

Received September 2015.

Contents

1. Introduction

Consider the logistic regression set-up in which $Y_1, \dots, Y_N$ are independent Bernoulli random variables such that $\Pr(Y_i = 1) = F(x_i^T \beta)$, where x_i is a $p \times 1$ vector of known covariates that are associated with Y_i, β is a $p \times 1$ vector of unknown regression coefficients, and F is the standard logistic distribution function, that is, $F(s) = e^s/(1+e^s)$. An important special case is the one-way logistic analysis of variance (ANOVA) model, where each x_i is a unit vector. (See Section 3 for a detailed explanation of how the logistic regression model reduces to the one-way model.) In general, the joint mass function of $Y_1, \dots, Y_N$ is given by

$$\prod_{i=1}^{N} \Pr(Y_i = y_i \,|\, \beta) = \prod_{i=1}^{N} \left[F(x_i^T\beta)\right]^{y_i} \left[1 - F(x_i^T\beta)\right]^{1-y_i} I_{\{0,1\}}(y_i)\,. \tag{1}$$

A Bayesian version of the logistic regression model (1) can be assembled by specifying a prior for the unknown regression parameter β. We consider a flat improper prior. Of course, whenever an improper prior is used, one must check that the resulting posterior is well-defined (i.e., proper). Chen and Shao (2000) provide necessary and sufficient conditions for propriety and these are stated explicitly in Section 2. We assume that these conditions are satisfied, and denote the posterior by $\pi(\beta \,|\, y)$. This posterior density is intractable in the sense that its expectations, which are required for Bayesian inference, cannot be computed in closed form. However, we may resort to Markov chain Monte Carlo (MCMC) methods to approximate the intractable posterior expectations.

The most recent MCMC method for this problem is a data augmentation (DA) algorithm that is based on the Polya-Gamma latent data strategy developed in Polson, Scott and Windle (2013) (hereafter, PS&W). In this article, we show that in the one-way logistic ANOVA model, which is an important special case of logistic regression models, the Markov operator associated with PS&W's DA algorithm is trace-class (see Section 3 for definition). The fact that this Markov operator is trace-class implies that it is also compact, which in turn implies that the corresponding Markov chain is *geometrically ergodic*. This is very important from a practical standpoint because geometric ergodicity guarantees the existence of central limit theorems for ergodic averages, which allows for the calculation of asymptotically valid standard errors for the MCMC estimates of posterior expectations (see, e.g., Flegal, Haran and Jones, 2008; Jones, Haran, Caffo and Neath, 2006).

Aside from our work, the only existing convergence rate result of a DA algorithm for the Bayesian logistic model is the one in Choi and Hobert (2013). However, in Choi and Hobert (2013), the model that was considered has proper normal priors on the regression parameter. While our result applies to a relatively small class of logistic regression models, it is the first of its kind for logistic regression models with an *improper* prior. It turns out that using a flat improper prior complicates the analysis that is required to study the corresponding Markov chain. Indeed, our analysis is substantially different from theirs.

The remainder of this paper is organized as follows. Section 2 contains a formal description of PS&W's algorithm for exploring the posterior $\pi(\beta \,|\, y)$. In Section 3, we study the operator associated with PS&W's DA Markov chain for the one-way logistic ANOVA model and show that the associated Markov operator is trace-class. Finally, in Section 4 we discuss possible difficulties in extending our result to the general logistic regression model.

2. Polson, Scott and Windle's algorithm

We begin with a description of the posterior density and conditions for propriety. We consider the posterior density that results when the logistic regression likelihood (1) is combined with a flat prior on the regression parameter β. Let $y = (y_1, \ldots, y_N)^T$ denote the vector of observed data and define the marginal density as

$$c(y) = \int_{\mathbb{R}^p} \prod_{i=1}^{N} \left[F(x_i^T\beta)\right]^{y_i} \left[1 - F(x_i^T\beta)\right]^{1-y_i} d\beta \ .$$

By definition, the posterior is proper if and only if $c(y) < \infty$. Of course, when $c(y) < \infty$, the posterior density is given by

$$\pi(\beta \,|\, y) = \frac{1}{c(y)} \prod_{i=1}^{N} \left[F(x_i^T\beta)\right]^{y_i} \left[1 - F(x_i^T\beta)\right]^{1-y_i} \ .$$

Recall from the Introduction that Chen and Shao (2000) provide necessary and sufficient conditions for propriety. In order to state Chen and Shao's (2000) result, we need a bit more notation. As usual, let X denote the $N \times p$ design matrix whose ith row is x_i^T, and let Z be an $N \times p$ matrix whose ith row is $z_i^T := I_{\{0\}}(y_i)\, x_i^T - I_{\{1\}}(y_i)\, x_i^T$. Finally, let 0_p be the $p \times 1$ vector of zeros. Here is the result.

Proposition 1 (Chen and Shao, 2000)**.** *The function $c(y)$ is finite if and only if*

(A) the design matrix X has full column rank and
(B) there is a vector $b = (b_1, \ldots, b_N)^T$ with strictly positive components such that $Z^T b = 0_p$.

In particular, in the one-way logistic ANOVA model, the design matrix X has full column rank so the posterior is proper if and only if (B) holds. (The precise form of X in the one-way model and an easily checkable equivalent condition for propriety are stated in the next section.) Throughout the remainder of this section, we assume that the conditions of Proposition 1 are satisfied so that the posterior is well-defined.

We now describe PS&W's DA algorithm for exploring the posterior $\pi(\beta \,|\, y)$. Let $\mathbb{R}_+ := (0, \infty)$. For fixed $w \in \mathbb{R}_+^N$, define $\Sigma(w) = (Z^T\Omega(w)Z)^{-1}$ and $\mu(w) = \Sigma(w)Z^T(-\frac{1}{2}1_N)$, where $\Omega(w)$ is the $N \times N$ diagonal matrix whose ith diagonal entry is w_i, and 1_N is the $N \times 1$ vector of 1's. When we write $W \sim \text{PG}(1, c)$, we mean W has a Polya-Gamma density (PS&W) given by

$$f(x; c) = \cosh(c/2)\, e^{-\frac{c^2 x}{2}}\, g(x) \ ,$$

where $c \geq 0$ and

$$g(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(2k+1)}{\sqrt{2\pi x^3}} e^{-\frac{(2k+1)^2}{8x}} I_{(0,\infty)}(x) \; . \tag{2}$$

The dynamics of PS&W's Markov chain, $\Phi = \{\beta^{(m)}\}_{m=0}^{\infty}$, are implicitly defined through the following two-step procedure for moving from the current state, $\beta^{(m)} = \beta$, to new state $\beta^{(m+1)}$.

Iteration $m+1$ of PS&W's DA algorithm:

1. Draw $W_1, \dots, W_N$ independently with
$$W_i \sim \mathrm{PG}(1, |z_i^T \beta|) \; ,$$
and call the observed value $w = (w_1, \dots, w_N)^T$.
2. Draw $\beta^{(m+1)} \sim \mathrm{N}_p\big(\mu(w), \Sigma(w)\big)$.

A highly efficient rejection sampler for simulating the Polya-Gamma distribution is provided in PS&W. Also, a formal derivation of the above algorithm is similar to the one for the normal prior case provided in Choi and Hobert (2013).

The Markov transition density (Mtd) of the Markov chain Φ is given by

$$k(\beta \,|\, \beta') = \int_{\mathbb{R}_+^N} \pi(\beta \,|\, w, y)\, \pi(w \,|\, \beta', y)\, dw \; ,$$

where $\pi(\beta \,|\, w, y)$ and $\pi(w \,|\, \beta, y)$ are conditional densities that can be gleaned from the two-step algorithm described above. Indeed, $\pi(w \,|\, \beta, y)$ is a product of Polya-Gamma densities, and $\pi(\beta \,|\, w, y)$ is the multivariate normal density. Note that k maps $\mathbb{R}^p \times \mathbb{R}^p$ into $\mathbb{R}_+$ and that $\pi(\beta \,|\, y)$ is an invariant density for this Mtd. It follows that the corresponding Markov chain Φ is Harris ergodic; that is, irreducible, aperiodic, and positive Harris recurrent (see, e.g., Hobert, 2011).

3. Main result

In this section, we restrict attention to the one-way logistic ANOVA model, an important special case of logistic regression models, and present a spectral analysis result concerning PS&W's Markov chain Φ for this problem. In particular, we prove that the Markov operator associated with Φ is trace-class.

We begin by describing the one-way logistic ANOVA model. Let $\{Y_{ij}\}$ be independent Bernoulli random variables such that

$$\Pr(Y_{ij} = 1 \,|\, \beta) = F(\beta_i) \; , \qquad i = 1, 2, \dots, p \; , \; j = 1, 2, \dots, n_i \; ,$$

where β_i is the main effect of the ith treatment group, and $\beta = (\beta_1, \ldots, \beta_p)^T$. Note that there are p treatment groups and the number of observations in different groups may differ. Of course, this model can be written in the logistic regression form (1). Indeed, there are a total of $N := n_1 + \cdots + n_p$ observations and we arrange them using the usual lexicographical ordering:

$$y = [y_{11} \cdots y_{1n_1} \quad y_{21} \cdots y_{2n_2} \quad \cdots \quad y_{p1} \cdots y_{pn_p}]^T .$$

The random version of y, call it Y, is defined similarly using the same ordering. We denote the lth components of y and Y as y_l and Y_l, respectively. Letting $\tilde{n}_0 = 0$ and $\tilde{n}_j = \sum_{k=1}^{j} n_k$ for $j \in \mathbb{N}_p := \{1, 2, \ldots, p\}$, we can write

$$x_i = e_j \ , \quad \text{if } \tilde{n}_{j-1} + 1 \le i \le \tilde{n}_j \ ,$$

where e_j denote the $p \times 1$ jth unit vector (i.e., the jth column of the $p \times p$ identity matrix).

Note that the design matrix X has full column rank since it equals $\bigoplus_{i=1}^{p} 1_{n_i}$, which has orthogonal columns. It follows from Proposition 1 that the posterior $\pi(\beta \,|\, y)$ that results when the one-way logistic ANOVA likelihood is combined with a flat prior on β is proper if and only if (B) holds, but this condition is not easy to interpret. We present an equivalent condition that is easy to check and understand. As usual, let $\hat{p}_i$ be the proportion of 1's in the ith treatment group, that is, $\hat{p}_i = \frac{1}{n_i} \sum_{l=\tilde{n}_{i-1}+1}^{\tilde{n}_i} y_l$. The following result, which is proven in the Appendix, implies that the posterior is proper if and only if there are at least one 1 and one 0 in *each* treatment group.

Corollary 1. *Assume that $X = \bigoplus_{i=1}^{p} 1_{n_i}$. The posterior is proper if and only if*

$$0 < \hat{p}_i < 1 \qquad \text{for all } i \in \mathbb{N}_p \ . \tag{B$'$}$$

Assume that the posterior $\pi(\beta \,|\, y)$ is proper. We now study PS&W's Markov chain Φ for exploring the posterior. In order to formally describe our main result, we must introduce the operator associated with Φ. Recall that $k(\beta \,|\, \beta')$ denotes the Mtd of Φ, that is, the conditional density of $\beta^{(m+1)}$ given that $\beta^{(m)} = \beta'$. Let L_0^2 be the space of real-valued functions with domain $\mathbb{R}^p$ that are square integrable and have mean zero with respect to the posterior density $\pi(\beta \,|\, y)$. This is a Hilbert space in which the inner product of $g, h \in L_0^2$ is defined as

$$\langle g, h \rangle = \int_{\mathbb{R}^p} g(\beta)\, h(\beta)\, \pi(\beta \,|\, y)\, d\beta \ ,$$

and the corresponding norm is, of course, given by $\|g\| = \langle g, g \rangle^{1/2}$. The Mtd k defines an operator on L_0^2 and the spectrum of the operator contains a great deal of information about the convergence behavior of the corresponding Markov chain Φ (see, e.g., Hobert, Roy and Robert, 2011). Let $K : L_0^2 \to L_0^2$ denote the operator that maps $g \in L_0^2$ to

$$(Kg)(\beta') = \int_{\mathbb{R}^p} g(\beta)\, k(\beta \,|\, \beta')\, d\beta \ .$$

Because the operator K is self-adjoint and positive, the spectrum of K is a subset of the interval $[0,1]$ (see, e.g., Hobert and Marchev, 2008; Liu, Wong and Kong, 1994; Hobert et al., 2011). Moreover, if the self-adjoint, positive operator K is compact, then its spectrum consists solely of eigenvalues (which are all strictly less than one) and the point $\{0\}$ (see, e.g., Retherford, 1993, p. 61-62). If the sum of the eigenvalues is finite, then the operator is called *trace-class* (see Khare and Hobert (2011), and references therein). Here is our main result.

Theorem 1. *Assume that* $X = \bigoplus_{i=1}^{p} 1_{n_i}$ *and the posterior is proper. Then, the Markov operator* K *is trace-class.*

The following lemma, which is a simple derivation of results in Devroye (2009), will be used in the proof of Theorem 1.

Lemma 1. *Let* g *denote the density of* $PG(1,0)$ *random variable. If* $w \in (0, \frac{1}{\log 3}]$, *then*

$$g(w) \leq \frac{1}{\sqrt{2\pi w^3}} \exp\left\{-\frac{1}{8w}\right\} .$$

Remark 1. *It is clear that (2) is a density of* $PG(1,0)$ *random variable. As described in Devroye (2009), Lemma 1 follows from the fact that* g *is of the form* $g(w) = \sum_{k=0}^{\infty}(-1)^k a_k(w)$ *where nonnegative* $\{a_k(w)\}_{k=0}^{\infty}$ *is decreasing in* k *for* $w \in (0, \frac{1}{\log 3}]$.

Proof of Theorem 1. To prove that the Markov operator K is trace-class, we follow a technique used in Khare and Hobert (2011); that is, we will establish the following condition

$$\int_{\mathbb{R}^p} k(\beta \,|\, \beta)\, d\beta = \int_{\mathbb{R}_+^N} \left[\int_{\mathbb{R}^p} \pi(\beta \,|\, w, y)\, \pi(w \,|\, \beta, y)\, d\beta\right] dw < \infty\, . \tag{3}$$

The key is to bound the inner integral in (3) by

$$c_1 \prod_{i=1}^{N} \left[\ \exp\ \left\{\frac{a}{8w_i}\right\} g(w_i)\right]\ ,$$

where c_1 and $a < 1$ are constants. We then use Lemma 1 to complete the proof.

We begin by evaluating the inner integral in (3). Recall that $\Sigma = \Sigma(w) = (Z^T\Omega(w)Z)^{-1}$ and $\mu = \mu(w) = \Sigma(w)Z^T(-\frac{1}{2}1_N)$. First, note that the product of densities $\pi(\beta \,|\, w, y) \times \pi(w \,|\, \beta, y)$ can be written as follows:

$$\begin{aligned}
&(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\beta^T(Z^T\Omega Z)\beta - 2\beta^T Z^T\Big(-\frac{1}{2}1_N\Big)\right)\right\}\\
&\quad\times \exp\left\{-\frac{1}{8}1_N^T Z(Z^T\Omega Z)^{-1}Z^T 1_N\right\}\\
&\quad\times \prod_{i=1}^{N} \cosh\left(\frac{|z_i^T\beta|}{2}\right) \exp\left\{-\frac{(z_i^T\beta)^2}{2}w_i\right\} g(w_i)
\end{aligned}$$

$$
\begin{aligned}
&= (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\beta^T\left(2\Sigma^{-1}\right)\beta\right\} \\
&\quad \times \prod_{i=1}^{N} \cosh\left(\frac{|z_i^T\beta|}{2}\right) \exp\left\{-\frac{z_i^T\beta}{2}\right\} g(w_i) \\
&\quad \times \exp\left\{-\frac{1}{8} 1_N^T Z (Z^T\Omega Z)^{-1} Z^T 1_N\right\} \\
&= (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\beta^T\left(2\Sigma^{-1}\right)\beta\right\} \left[\prod_{i=1}^{N} \frac{1}{2}\left(1+\exp\{-z_i^T\beta\}\right)\right] \\
&\quad \times \exp\left\{-\frac{1}{8} 1_N^T Z (Z^T\Omega Z)^{-1} Z^T 1_N\right\} \prod_{i=1}^{N} g(w_i) \;, \qquad (4)
\end{aligned}
$$

where the last equality follows from $\cosh(a)e^{-a} = \frac{1}{2}(1+e^{-2a})$.

We now evaluate the integral of (4) with respect to β. Recall $\mathbb{N}_N = \{1, 2, \ldots, N\}$. For each $A \subseteq \mathbb{N}_N$, define an $N \times p$ matrix Z_A whose ith row is

$$
\begin{cases} z_i^T & \text{if } i \in A \\ 0_p^T & \text{if } i \in \mathbb{N}_N \setminus A \,. \end{cases}
$$

Then, it is easy to see that

$$
\exp\{-1_N^T Z_A \beta\} = \begin{cases} \exp\left\{-\sum_{i\in A} z_i^T\beta\right\} & \text{if } A \text{ is nonempty} \\ 1 & \text{if } A \text{ is empty} \,. \end{cases}
$$

Therefore, we have

$$
\prod_{i=1}^{N}\left(1+\exp\{-z_i^T\beta\}\right) = \sum_{A\subseteq\mathbb{N}_N} \exp\left\{-1_N^T Z_A\beta\right\}
$$

so the integral of (4) with respect to β can be written as

$$
\begin{aligned}
&c_0 \exp\left\{-\frac{1}{8} 1_N^T Z (Z^T\Omega Z)^{-1} Z^T 1_N\right\} \left[\prod_{i=1}^{N} g(w_i)\right] \\
&\quad \times \sum_{A\subseteq\mathbb{N}_N} \int_{\mathbb{R}^p} \exp\left\{-1_N^T Z_A\beta\right\} \phi\left(\beta;\, 0_p, \Sigma/2\right) d\beta \;, \qquad (5)
\end{aligned}
$$

where c_0 is a constant, and $\phi(\beta;\, 0_p, \Sigma/2)$ is a multivariate normal density with mean 0_p and variance $\Sigma/2$. Note that the integral in (5) is just the moment generating function of β evaluated at the point $-1_N^T Z_A$. Hence, (5) is equal to

$$
c_0 \exp\left\{-\frac{1}{8} 1_N^T Z (Z^T\Omega Z)^{-1} Z^T 1_N\right\} \left[\prod_{i=1}^{N} g(w_i)\right]
$$

$$\times \sum_{A \subseteq \mathbb{N}_N} \exp\left\{ \frac{1}{4} 1_N^T Z_A (Z^T \Omega Z)^{-1} Z_A^T 1_N \right\} . \tag{6}$$

We now express the exponential terms of (6) in a more compact way. Recall $X = \bigoplus_{i=1}^p 1_{n_i}$ and the ith row of Z is $z_i^T = I_{\{0\}}(y_i)\, x_i^T - I_{\{1\}}(y_i)\, x_i^T$. Clearly, the $N \times p$ matrix Z can be written as UX, where U is the $N \times N$ diagonal matrix whose ith diagonal entry is equal to $u_i := I_{\{0\}}(y_i) - I_{\{1\}}(y_i)$. It is easy to see $Z^T \Omega Z = X^T \Omega X$, and using the orthogonality of the columns of X, $Z^T \Omega Z$ is a $p \times p$ diagonal matrix whose ith diagonal entry is equal to

$$\sum_{j=\tilde{n}_{i-1}+1}^{\tilde{n}_i} w_j \ ,$$

where $\tilde{n}_0 = 0$ and $\tilde{n}_i = \sum_{k=1}^i n_k$ for $i \in \mathbb{N}_p$. Let z_{ij} denote the jth component of the row vector z_i^T. A straightforward calculation yields that, for $A \subseteq \mathbb{N}_N$,

$$1_N^T Z_A (Z^T \Omega Z)^{-1} Z_A^T 1_N = \sum_{j=1}^p \frac{\left(\sum_{i \in A} z_{ij}\right)^2}{\sum_{k=\tilde{n}_{j-1}+1}^{\tilde{n}_j} w_k} \ .$$

Hence, the exponential terms in (6) can be written as

$$\sum_{A \subseteq \mathbb{N}_N} \exp\left\{ \frac{1}{8} \sum_{j=1}^p \frac{1}{\sum_{k=\tilde{n}_{j-1}+1}^{\tilde{n}_j} w_k} \left[2 \left(\sum_{i \in A} z_{ij} \right)^2 - \left(\sum_{i=1}^N z_{ij} \right)^2 \right] \right\} . \tag{7}$$

We now claim that, for all $A \subseteq \mathbb{N}_N$ and all $j \in \mathbb{N}_p$,

$$\frac{1}{n_j^2} \left[2 \left(\sum_{i \in A} z_{ij} \right)^2 - \left(\sum_{i=1}^N z_{ij} \right)^2 \right] < 1 \ .$$

Recall that $\hat{p}_j$ is the proportion of 1's in the jth treatment group, that is, $\hat{p}_j = \frac{1}{n_j} \sum_{i=\tilde{n}_{j-1}+1}^{\tilde{n}_j} y_i$. Also, recall from the definition of Z that its jth column is

$$\begin{bmatrix} 0_{\tilde{n}_{j-1}}^T & u_{\tilde{n}_{j-1}+1} & \cdots & u_{\tilde{n}_j} & 0_{N-\tilde{n}_j}^T \end{bmatrix}^T \ ,$$

where $u_i = I_{\{0\}}(y_i) - I_{\{1\}}(y_i)$. Of course, when $j = 1$ or $j = p$, we omit the vector 0_0; for example, the first column of Z is $[u_1 \ \cdots \ u_{n_1} \ 0_{N-n_1}^T]^T$. Since there are $n_j \hat{p}_j$ 1's and $n_j(1-\hat{p}_j)$ 0's in the jth treatment group, in the jth column of Z, $n_j \hat{p}_j$ of the u_i's are -1's and $n_j(1-\hat{p}_j)$ of the u_i's are 1's. Assume without loss of generality that, for each $j \in \mathbb{N}_p$, the first $n_j \hat{p}_j$ u_i's are -1's and the remaining $n_j(1-\hat{p}_j)$ u_i's are 1's. That is, we assume that the jth column of Z is equal to

$$\begin{cases} [\, -1_{n_1\hat{p}_1}^T \quad 1_{n_1(1-\hat{p}_1)}^T \quad 0_{N-n_1}^T \,]^T & \text{if } j = 1 \\ [\, 0_{\tilde{n}_{j-1}}^T \quad -1_{n_j\hat{p}_j}^T \quad 1_{n_j(1-\hat{p}_j)}^T \quad 0_{N-\tilde{n}_j}^T \,]^T & \text{if } 1 < j < p \\ [\, 0_{\tilde{n}_{p-1}}^T \quad -1_{n_p\hat{p}_p}^T \quad 1_{n_p(1-\hat{p}_p)}^T \,]^T & \text{if } j = p \ . \end{cases} \tag{8}$$

Letting $\hat{q}_j = 1 - \hat{p}_j$ for $j \in \mathbb{N}_p$, the sum of the jth column of Z is

$$\sum_{i=1}^{N} z_{ij} = n_j(1-\hat{p}_j) - n_j\hat{p}_j = n_j(\hat{q}_j - \hat{p}_j) \ ,$$

and, for all $A \subseteq \mathbb{N}_N$ and $j \in \mathbb{N}_p$,

$$\left| \sum_{i \in A} z_{ij} \right| \le \max\left\{ n_j\hat{p}_j, n_j\hat{q}_j \right\} \ .$$

It follows that, for all $A \subseteq \mathbb{N}_N$ and all $j \in \mathbb{N}_p$,

$$\begin{aligned} \frac{1}{n_j^2}\left[2\left(\sum_{i\in A} z_{ij}\right)^2 - \left(\sum_{i=1}^{N} z_{ij}\right)^2 \right] &\le 2\max\{\hat{p}_j, \hat{q}_j\}^2 - (\hat{q}_j - \hat{p}_j)^2 \\ &= 2(\hat{p}_j^2 + \hat{q}_j^2) - 2\min\{\hat{p}_j, \hat{q}_j\}^2 - (\hat{q}_j - \hat{p}_j)^2 \\ &= 1 - 2\min\{\hat{p}_j, \hat{q}_j\}^2 \ . \end{aligned}$$

Since condition (B′) of Corollary 1 is in force, for all $j \in \mathbb{N}_p$, we have

$$1 - 2\min\{\hat{p}_j, \hat{q}_j\}^2 < 1 \ .$$

Hence, (7) is bounded above by

$$2^N \exp\left\{ \frac{a}{8} \sum_{j=1}^{p} \frac{n_j^2}{\sum_{k=\tilde{n}_{j-1}}^{\tilde{n}_j} w_k} \right\} \ , \tag{9}$$

where

$$a := \max_{j \in \mathbb{N}_p} \left[1 - 2\min\{\hat{p}_j, \hat{q}_j\}^2 \right] < 1 \ .$$

Combining (5), (6), (7) and (9), we have

$$\begin{aligned} \int_{\mathbb{R}^p} \pi(\beta \,|\, w, y)\, \pi(w \,|\, \beta, y)\, d\beta &\le c_1 \exp\left\{ \frac{a}{8} \sum_{j=1}^{p} \frac{n_j^2}{\sum_{k=\tilde{n}_{j-1}+1}^{\tilde{n}_j} w_k} \right\} \prod_{i=1}^{N} g(w_i) \\ &\le c_1 \exp\left\{ \frac{a}{8} \sum_{j=1}^{p} \sum_{k=\tilde{n}_{j-1}+1}^{\tilde{n}_j} \frac{1}{w_k} \right\} \prod_{i=1}^{N} g(w_i) \\ &= c_1 \prod_{i=1}^{N} \exp\left\{ \frac{a}{8w_i} \right\} g(w_i) \ , \end{aligned}$$

where c_1 is a constant, and the last inequality is due to the arithmetic-harmonic mean inequality. It follows that

$$\int_{\mathbb{R}_+^N} \int_{\mathbb{R}^p} \pi(\beta \,|\, w, y)\, \pi(w \,|\, \beta, y)\, d\beta\, dw$$

$$
\begin{aligned}
&\leq c_1 \prod_{i=1}^{N} \int_{\mathbb{R}_+} \exp\left\{\frac{a}{8w_i}\right\} g(w_i)\, dw_i \\
&\leq c_1 \prod_{i=1}^{N} \left[\int_0^t \exp\left\{\frac{a}{8w_i}\right\} g(w_i)\, dw_i + \int_t^\infty \exp\left\{\frac{a}{8t}\right\} g(w_i)\, dw_i\right] \\
&\leq c_1 \prod_{i=1}^{N} \left[\int_0^t \frac{1}{\sqrt{2\pi w_i^3}} \exp\left\{-\frac{1-a}{8w_i}\right\} dw_i + c_2\right] , \qquad (10)
\end{aligned}
$$

where $t = \frac{1}{\log 3}$, c_2 is a constant, and the last inequality is due to Lemma 1. The integrand in (10) is a kernel of an inverse gamma density, and hence (3) is satisfied. □

4. Discussion

We have shown in Section 3 that in the one-way logistic ANOVA model, the Markov operator associated with PS&W's DA chain is trace-class (and thus the chain is geometrically ergodic). Given this result, it is natural to ask whether PS&W's DA operator is also trace-class in the general logistic regression model. We note that our trace-class proof for the one-way model relies heavily on the orthogonality of the columns of the design matrix X, which is not guaranteed in general. In particular, the idea is to use orthogonality to express $(Z^T\Omega Z)^{-1}$ as a diagonal matrix whose diagonal entries are simple functions of $w = (w_1, \ldots, w_N)^T$. However, in the general case, $(Z^T\Omega Z)^{-1}$ does not necessarily take a simple form, which in turn complicates the analysis. Thus, we believe that establishing (3) for the general case would be much more than a straightforward extension of the proof of our Theorem 1. It is our hope that our result for the one-way logistic ANOVA model will promote the study of the DA Markov chain for the general logistic regression model.

Appendix A: Appendix

In this section, we provide a proof of Corollary 1. Before presenting the proof, we recall the structure of Z given in (8) which will be used a couple of times within the proof. The jth column of the $N \times p$ matrix Z is

$$
\begin{cases}
[\,-1^T_{n_1\hat{p}_1} \quad 1^T_{n_1(1-\hat{p}_1)} \quad 0^T_{N-n_1}\,]^T & \text{if } j = 1 \\
[\,0^T_{\tilde{n}_{j-1}} \quad -1^T_{n_j\hat{p}_j} \quad 1^T_{n_j(1-\hat{p}_j)} \quad 0^T_{N-\tilde{n}_j}\,]^T & \text{if } 1 < j < p \\
[\,0^T_{\tilde{n}_{p-1}} \quad -1^T_{n_p\hat{p}_p} \quad 1^T_{n_p(1-\hat{p}_p)}\,]^T & \text{if } j = p\,.
\end{cases}
$$

Proof of Corollary 1. Recall that $c(y) < \infty$ if and only if (B) holds. Thus, it suffices to show that (B′) is equivalent to (B). First, we shall write $b = (b_1^T, \cdots, b_p^T)^T$, where each b_i is an $n_i \times 1$ vector. We start by demonstrating that (B) implies (B′). We will establish this by contradiction. Assume (without loss of generality) that $\hat{p}_1 = 0$. Then, there are no -1's in the first column of Z, so the first column of Z is equal to $[1_{n_1}^T \;\; 0_{N-n_1}^T]^T$. It follows that the first component of the $p \times 1$ vector $Z^T b$ is $1_{n_1}^T b_1 > 0$ for all $b \in \mathbb{R}_+^N$, which contradicts (B). We can similarly prove that the assumption of $\hat{p}_1 = 1$ yields a contradiction to (B). Hence, it must be that $0 < \hat{p}_j < 1$ for all $j \in \mathbb{N}_p$.

We now show that (B′) implies (B). We will establish this by explicitly constructing a vector $b \in \mathbb{R}_+^N$ such that $Z^T b = 0_p$. For each $j \in \mathbb{N}_p$, define

$$b_j^T = \begin{cases} [\, n_j(1-2\hat{p}_j)+1 \quad 1_{n_j-1}^T \,] & \text{if } 0 < \hat{p}_j < \frac{1}{2} \\ [\, 1_{n_j-1}^T \quad n_j(2\hat{p}_j-1)+1 \,] & \text{if } \frac{1}{2} \le \hat{p}_j < 1 \, . \end{cases}$$

Clearly, each b_j is in $\mathbb{R}_+^{n_j}$, so that $b \in \mathbb{R}_+^N$. Therefore, we need only to show that $Z^T b = 0_p$. It follows that, if $\frac{1}{2} \le \hat{p}_j < 1$, then the jth component of the $p \times 1$ vector $Z^T b$ is equal to

$$\begin{bmatrix} -1_{n_j \hat{p}_j}^T & 1_{n_j(1-\hat{p}_j)}^T \end{bmatrix} \begin{bmatrix} 1_{n_j-1} \\ n_j(2\hat{p}_j-1)+1 \end{bmatrix} = 0 \, .$$

We can similarly show that, if $0 < \hat{p}_j < \frac{1}{2}$, then the jth component of $Z^T b$ is also zero. Hence, $Z^T b = 0_p$, which completes the proof. $\square$

Acknowledgements

Román's research was supported by NSF Grant DMS-13-08765.

References

Chen, M.-H. and Shao, Q.-M. (2000). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*, **129** 293–302. MR1694452

Choi, H. M. and Hobert, J. P. (2013). The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics*, **7** 2054–2064. MR3091616

Devroye, L. (2009). On exact simulation algorithms for some distributions related to Jacobi theta functions. *Statistics and Probability Letters 79*, **79** 2251–2259.

Flegal, J. M., Haran, M. and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, **23** 250–260. MR2516823

Hobert, J. P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.). Chapman & Hall/CRC Press. MR2742422

Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *The Annals of Statistics*, **36** 532–554. MR2396806

Hobert, J. P., Roy, V. and Robert, C. P. (2011). Improving the convergence properties of the data augmentation algorithm with an application to Bayesian mixture modelling. *Statistical Science*, **26** 332–351.

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.

Khare, K. and Hobert, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics*, **39** 2585–2606.

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*, **81** 27–40.

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, **108** 1339–1349. MR3174712

Retherford, J. R. (1993). *Hilbert Space: Compact Operators and the Trace Theorem.* Cambridge University Press.