Geometric ergodicity of a random-walk Metropolis algorithm for a transformed density

By

Leif Johnson and Charles J. Geyer

Technical Report No. 680

School of Statistics

University of Minnesota

November 22, 2010

Abstract

Curvature conditions on a target density in $\mathbb{R}^k$ for the geometric ergodicity of a random-walk Metropolis algorithm have previously been established (Mengersen and Tweedie, 1996; Roberts and Tweedie, 1996; Jarner and Hansen, 2000). However, the conditions for target densities in $\mathbb{R}^k$ that have exponentially light tails, but are not super-exponential are difficult to apply. In this paper I establish a variable transformation to apply to such target densities, that along with a regularity condition on the target density, ensures that a random-walk Metropolis algorithm for the transformed density is geometrically ergodic. Inference can be drawn for the original target density using Markov chain Monte Carlo estimates based on the transformed density. An application to inference on the regression parameter in multinomial logit regression with a conjugate prior is given.

1 Introduction

Markov chain Monte Carlo (MCMC) is an important technique in modern statistics. Unless a Markov chain converges sufficiently fast the behavior and validity of MCMC estimates may not be reliable. Let π be a target density on $\mathbb{R}^k$, and $\mathrm{E}_\pi(g(\beta))$ be the expectation of g with respect to π. Then if a Markov chain with a stationary distribution specified by π is geometrically ergodic and either the chain is reversible or $\mathrm{E}_\pi(|g(\beta)|^{2+\varepsilon})$ exists for some $\varepsilon > 0$, and $\mathrm{E}_\pi(g(\beta)^2)$ is finite, then the MCMC estimate of $\mathrm{E}_\pi(g(\beta))$ will have a central limit theorem. Hence, establishing the geometric ergodicity of a Markov chain for a specific target density is important to drawing valid inference.

Let h be a 1-1 mapping. Then if the random variable β has density π_β and the random variable $\gamma = h(\beta)$ has density π_γ then $\mathrm{E}_{\pi_\beta}(g(\beta)) = \mathrm{E}_{\pi_\gamma}(g(h^{-1}(\gamma)))$. In this report we establish that for a class of densities we may not be able to find a geometrically ergodic Markov chain, but using a specific 1-1 mapping we can establish that a random-walk Metropolis algorithm is geometrically ergodic for the density of the transformed variable. Thus we can draw inference on the original target density by using a Markov chain on the density of the transformed variable. The posterior density of the regression parameter in a multinomial logit model is one such target density.

Section 2 reviews some of the previous curvature results, in particular the necessary results for our theorem. Section 3 defines the variable transformation used, and gives our main result. Section 4 shows that the result in Section 3 applies to the posterior density of the regression parameter in a multinomial logit model.

2 Curvature Conditions

A density π on $\mathbb{R}^k$ is said to have *exponentially light tails* if

$$-\infty < \limsup_{|x|\to\infty} \frac{x^T \nabla \log \pi(x)}{|x|} < 0. \tag{1}$$

Mengersen and Tweedie (1996) showed that if π is a continuous and always positive density on $\mathbb{R}$, then π having exponentially light tails is a necessary and sufficient condition for the geometric ergodicity of a random-walk Metropolis algorithm for π. Roberts and Tweedie

(1996) gave a complicated set of conditions for the geometric ergodicity of a random-walk Metropolis algorithm on a density in $\mathbb{R}^k$. A density π on $\mathbb{R}^k$ is *super-exponential* if it satisfies the following conditions

$$\pi \text{ is always positive and has continuous first derivatives} \tag{2}$$

$$\limsup_{|\beta|\to\infty} \frac{\beta^T \nabla \log \pi(\beta)}{|\beta|} = -\infty. \tag{3}$$

To be a random-walk, the proposal density, for x and x' q must satisfy the condition $q(x, x') = q(|x - x'|)$. We will need a further condition on q, that for some $\varepsilon_q > 0$ and $\delta_q > 0$.

$$q(x) \geq \varepsilon_q \text{ for } |x| \leq \delta_q. \tag{4}$$

Jarner and Hansen (2000) give the following Theorem for a density on $\mathbb{R}^k$,

Theorem 1. *If π is a super-exponential density on $\mathbb{R}^k$ and π satisfies*

$$\limsup_{|\beta|\to\infty} \frac{\beta^T}{|\beta|} \frac{\nabla\pi(\beta)}{|\nabla\pi(\beta)|} < 0, \tag{5}$$

then the random-walk-based Metropolis algorithm with q satisfying (4) *is geometrically ergodic.*

3 Variable Transformation Method

3.1 Variable Transformation

If $f : \mathbb{R}^k \mapsto \mathbb{R}$ is the density of x, and $h^{-1} : \mathbb{R}^k \mapsto \mathbb{R}^k$ is a $\mathcal{C}^1$ diffeomorphism, where $y = h(x)$, then the density of y is given by the variable transformation formula

$$f_y(y) = f(h^{-1}(y))|\nabla h^{-1}(y)|$$

We will now label target densities by the random variable they refer to, π_β will refer to the density for β and π_γ will refer to the density for γ. The goal is to define a transformation, h such that for $\gamma = h(\beta)$ that the density π_γ is super-exponential even though π_β has exponentially light tails but is not super-exponential. Conditions (2) will require that h^{-1} satisfy the conditions

$$h^{-1} \text{ is 1-1 and onto} \tag{6}$$

$$\gamma \mapsto |\nabla_\gamma h^{-1}(\gamma)| \text{ is continuous and always positive} \tag{7}$$

$$\gamma \mapsto \nabla_\gamma |\nabla_\gamma h^{-1}(\gamma)| \text{ is continuous.} \tag{8}$$

First define

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

then we can define $g : \mathbb{R} \mapsto \mathbb{R}$ as follows

$$g(x) = \begin{cases} \operatorname{sign}(x)(x^2 + 1/3) & \text{if } |x| > 1 \\ \frac{1}{3}x^3 + x & \text{if } |x| \leq 1 \end{cases} \tag{9}$$

which has first and second derivatives

$$g'(x) = \begin{cases} 2\operatorname{sign}(x)x & \text{if } |x| > 1 \\ x^2 + 1 & \text{if } |x| \leq 1 \end{cases} \tag{10}$$

$$g''(x) = \begin{cases} 2\operatorname{sign}(x) & \text{if } |x| > 1 \\ 2x & \text{if } |x| \leq 1. \end{cases} \tag{11}$$

Note that g is 1-1 and onto, g' is continuous and always positive, and g'' is continuous.

Now we can define $h^{-1} : \mathbb{R}^k \mapsto \mathbb{R}^k$ component-wise,

$$h^{-1}(\gamma)_j = g(\gamma_j). \tag{12}$$

We can also define the matrix $\nabla_\gamma h^{-1}(\gamma)$ component-wise

$$\nabla_\gamma h^{-1}(\gamma)_{ij} = \begin{cases} g'(\gamma_j) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{13}$$

Hence $\nabla_\gamma h^{-1}(\gamma)$ is a square diagonal matrix. So

$$|\nabla_\gamma h^{-1}(\gamma)| = \prod_j g'(\gamma_j). \tag{14}$$

Now we can write $\nabla_\gamma |\nabla_\gamma h^{-1}(\gamma)|$ component-wise by

$$\big(\nabla_\gamma |\nabla_\gamma h^{-1}(\gamma)|\big)_j = \frac{g''(\gamma_j)}{g'(\gamma_j)} |\nabla_\gamma h^{-1}(\gamma)|. \tag{15}$$

Since g is 1-1 and onto, h^{-1} satisfies condition (6). Since g' is continuous and always positive, $|\nabla_\gamma h^{-1}(\gamma)|$ satisfies condition (7). Since g' is always positive and continuous, and g'' is continuous $\nabla_\gamma |\nabla_\gamma h^{-1}(\gamma)|$ satisfies condition (8).

3.2 Supporting Lemmas

This section will give some lemmas that we will need to prove our theorems.

Lemma 1. *If π_β is a density with exponentially light tails, and h^{-1} is defined as in* (12)*, then*

$$\lim_{|\gamma| \to \infty} |\nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)| = \infty. \tag{16}$$

Proof. Equation (16) holding is equivalent to

$$\lim_{|\gamma| \to \infty} |\nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)|^2 = \infty. \tag{17}$$

holding. Let $Z(\gamma)$ be the set $\{\, j \mid g(\gamma_j) \neq 0 \,\}$. Then squared norm in (17) is bounded below by

$$\sum_{j \in Z(\gamma)} g'(\gamma_j)^2 \big(\nabla_\beta \log \pi_\beta(h^{-1}(\gamma))\big)_j^2$$

which can be rewritten as

$$\sum_{j \in Z(\gamma)} |g(\gamma_j)|^2 \big(\nabla_\beta \log \pi_\beta(h^{-1}(\gamma))\big)_j^2 / |g(\gamma_j)| \; \frac{g'(\gamma_j)^2}{|g(\gamma_j)|}. \tag{18}$$

Note that we are not dividing by 0 because we are limiting ourselves to $j \in Z(\gamma)$. For all γ, the fraction in (18) is bounded below by 2. The actual greatest lower bound can be shown to be approximately 2.72, but all we need is that the bound is positive. Noting that $|g(\gamma_j)|$ is bounded above by $|h^{-1}(\gamma)|$, we can multiply and divide by $|h^{-1}(\gamma)|$ (assuming that $|\gamma|$ is not 0) and bound (18) below by

$$2|h^{-1}(\gamma)| \sum_j \left(\frac{g(\gamma_j) \nabla_\beta \log \pi_\beta(h^{-1}(\gamma))}{|h^{-1}(\gamma)|} \right)^2. \tag{19}$$

Switching the indices in the sum from $j \in Z(\gamma)$ to all j is accomplished by adding 0, or the sum across $j \notin Z(\gamma)$. Since π_β has exponentially light tails, if we apply the triangle inequality, it is clear that for large $|\gamma|$ that there exists a $\varepsilon > 0$ such that

$$\sum_j \left| \frac{g(\gamma_j) \nabla_\beta \log \pi_\beta(h^{-1}(\gamma))}{|h^{-1}(\gamma)|} \right| \geq \varepsilon.$$

Therefore, there must exist some ε' such that for large enough $|\gamma|$, (19) is bounded below by $2|h^{-1}(\gamma)|\varepsilon'$. Therefore the limit in (17) is bounded below by

$$\liminf_{|\gamma| \to \infty} 2|h^{-1}(\gamma)|\varepsilon' = \infty.$$

thus equation (16) must hold. □

3.3 Transform an exponentially light density to super exponential

Theorem 2. *If π_β has exponentially light tails, is always positive and has continuous first derivatives,*

$$0 < \limsup_{|\beta| \to \infty} |\nabla_\beta \log \pi_\beta(\beta)| < \infty \tag{20}$$

and $\gamma = h(\beta)$ for h^{-1} defined in (12)*, then π_γ is a super-exponential density.*

Proof. We can define π_γ with the variable transformation formula

$$\pi_\gamma(\gamma) = \pi_\beta(h^{-1}(\gamma)) |\nabla_\gamma h^{-1}(\gamma)| \tag{21}$$

Since both π_β and $|\nabla_\gamma h^{-1}(\gamma)|$ are always positive functions with continuous first derivatives, π_γ is also an always positive function with continuous first derivatives. Hence π_γ satisfies condition (2). Now,

$$\nabla_\gamma \log \pi_\gamma(\gamma) = \nabla_\gamma h^{-1}(\gamma) \big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma) + \nabla_\gamma \log |\nabla_\gamma h^{-1}(\gamma)|. \tag{22}$$

Therefore, if we replace π with π_β in (3), then the lim sup can be bounded above by the sum of the lim sups of the two following terms

$$\frac{\gamma^T \nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)}{|\gamma|} \tag{23}$$

$$\frac{\gamma^T \nabla_\gamma \log |\nabla_\gamma h^{-1}(\gamma)|}{|\gamma|} \tag{24}$$

The numerator in (24) can be rewritten as

$$\sum_j \gamma_j \frac{g''(\gamma_j)}{g'(\gamma_j)}.$$

The definitions (10) and (11) tells us that

$$\gamma_j \frac{g''(\gamma_j)}{g'(\gamma_j)} = \begin{cases} 1 & \text{if } |\gamma_j| > 1 \\ \frac{2\gamma_j^2}{\gamma_j^2+1} & \text{if } |\gamma_j| \leq 1 \end{cases}$$

which is clearly bounded above by 1 for all γ_j. Therefore the numerator in (24) is clearly bounded as $|\gamma|$ goes to ∞, hence the lim sup as $|\gamma|$ goes to ∞ of (24) equals 0. All that remains to be shown is that the lim sup of (23) as $|\gamma|$ goes to ∞ is $-\infty$.

By multiplying and dividing by 2, and adding and subtracting $h^{-1}(\gamma)$, (23) may be written as the sum of the following two terms

$$2\frac{h^{-1}(\gamma)^T\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)}{|\gamma|} \tag{25}$$

$$2\frac{\big(\nabla_\gamma h^{-1}(\gamma)\gamma/2 - h^{-1}(\gamma)\big)^T\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)}{|\gamma|} \tag{26}$$

If we set

$$n_1(\gamma) = |\nabla_\gamma h^{-1}(\gamma)\gamma/2 - h^{-1}(\gamma)|$$
$$n_2(\gamma) = |\big(\nabla_\beta \log \pi_\beta\big) h^{-1}(\gamma)|$$

the Cauchy-Schwarz inequality tells us that the absolute value of (26) is bounded above by

$$2n_1(\gamma)n_2(\gamma)/|\gamma|.$$

Clearly

$$n_1(\gamma)^2 = \sum_j \big(\gamma_j g'(\gamma_j)/2 - g(\gamma_j)\big)^2.$$

Definitions (9) and (10) tell us that

$$\gamma_j g'(\gamma_j)/2 - g(\gamma_j) = \begin{cases} \mathrm{sign}(\gamma_j)\gamma_j^2 - \mathrm{sign}(\gamma_j)(\gamma_j^2 + 1/3) & \text{if } |\gamma_j| > 1 \\ \gamma_j^3/2 + \gamma_j/2 - \gamma_j^3/3 - \gamma_j & \text{if } |\gamma_j| \leq 1 \end{cases}$$

which is clearly bounded for all γ_j, hence $n_1(\gamma)$ is bounded for all γ. Now, for $|\gamma|$ large enough, our assumptions state that $n_2(\gamma)$ is bounded. Hence for $|\gamma|$ large the numerator in (26) is bounded, hence the lim sup as $|\gamma|$ goes to ∞ of (26) is 0. Therefore the lim sup in (3) is bounded above by the lim sup of (25).

Equation (25) can be rewritten as

$$2\frac{|h^{-1}(\gamma)|}{|\gamma|}\frac{h^{-1}(\gamma)^T\big(\nabla_\beta \log \pi_\beta\big)(h^{-1}(\gamma))}{|h^{-1}(\gamma)|}.$$

Our assumptions tell us that for $|\gamma|$ large enough, and some $\varepsilon > 0$

$$\frac{h^{-1}(\gamma)^T\big(\nabla_\beta \log \pi_\beta\big)(h^{-1}(\gamma))}{|h^{-1}(\gamma)|} \leq -\varepsilon$$

Hence, for $|\gamma|$ large enough, (25) is bounded above by

$$-2\varepsilon\frac{|h^{-1}(\gamma)|}{|\gamma|}. \tag{27}$$

Clearly as $|\gamma|$ goes to ∞,

$$|h^{-1}(\gamma)| \geq \max_j \sqrt{(\gamma_j^2 + 1/3)^2} \geq \max_j \gamma_j^2$$

and

$$|\gamma| \leq \sqrt{m}\max_j |\gamma_j|$$

and $\arg\max_j |\gamma_j| = \arg\max_j \gamma_j^2$. Therefore (27) is bounded above by

$$-2\varepsilon\max_j |\gamma_j|$$

which clearly goes to $-\infty$ as $|\gamma|$ goes to ∞. Hence the lim sup of (25) is $-\infty$ so the lim sup in (3), so π_γ satisfies (3). □

Lemma 2. *Let π_β is an always positive density with continuous first derivatives, and $|\nabla_\beta \log \pi_\beta(\beta)|$ be bounded away from 0 and ∞ as $|\beta|$ goes to ∞. Then if h^{-1} is defined as in* (12) *and $\gamma = h(\beta)$, then*

$$\lim_{|\gamma|\to\infty}\frac{|\nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)|}{|\nabla_\gamma \log \pi_\gamma(\gamma)|} = 1 \tag{28}$$

Proof. Set

$$n_1(\gamma) = \nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)$$

and recall Lemma 1 shows that

$$\lim_{|\gamma|\to\infty}|n_1(\gamma)| = \infty.$$

Since both the numerator and denominator in (28) are non-negative numbers, (28) is equivalent to

$$\lim_{|\gamma|\to\infty}\frac{|\nabla_\gamma \log \pi_\gamma(\gamma)|^2}{|n_1(\gamma)|^2} = 1. \tag{29}$$

The fraction in (29) can be rewritten as the sum of three terms

$$|n_1(\gamma)|^2 \;/\; |n_1(\gamma)|^2 \tag{30a}$$

$$2\big(\nabla_\gamma \log|\nabla_\gamma h^{-1}(\gamma)|\big)^T n_1(\gamma) \;/\; |n_1(\gamma)|^2 \tag{30b}$$

$$|\nabla_\gamma \log|\nabla_\gamma h^{-1}(\gamma)|^2 \;/\; |n_1(\gamma)|^2 \tag{30c}$$

We already know that numerator of (30c) is bounded and the denominator is unbounded as $|\gamma|$ goes to ∞, so (30c) goes to 0 as $|\gamma|$ goes to ∞.

The Cauchy-Schwarz inequality bounds the absolute value of (30b) above by

$$2\frac{\big|\nabla_\gamma \log|\nabla_\gamma h^{-1}(\gamma)|\big|}{|n_1(\gamma)|}.$$

The numerator of this fraction is bounded and the denominator is unbounded as $|\gamma|$ goes to ∞, hence (30b) goes to 0 as $|\gamma|$ goes to ∞.

Therefore the fraction in (29) is the sum of three terms, 1 plus two terms that go to 0 as $|\gamma|$ goes to ∞, hence (29) holds and therefore (28) holds. □

Theorem 3. *If π_β has exponentially light tails, is always positive and has continuous first derivatives,*

$$\limsup_{|\beta|\to\infty} |\nabla_\beta \log \pi_\beta(\beta)| < \infty \tag{31}$$

$$0 < \liminf_{|\beta|\to\infty} |\nabla_\beta \log \pi_\beta(\beta)| \tag{32}$$

and $\gamma = h(\beta)$ for h^{-1} defined in (12)*, then π_γ satisfies condition* (5)*.*

Proof. Since the conditions of Theorem 2 are satisfied π_γ is always positive, so $\nabla_\gamma \pi_\gamma(\gamma) = \pi_\gamma(\gamma)\nabla_\gamma \log \pi_\gamma(\gamma)$ and we can replace condition (5) with

$$\limsup_{|\gamma|\to\infty} \frac{\gamma^T \nabla_\gamma \log \pi_\gamma(\gamma)}{|\gamma||\nabla_\gamma \log \pi_\gamma(\gamma)|} < 0. \tag{33}$$

The fraction in (33) can be rewritten as the sum

$$\frac{\gamma^T \nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)}{|\gamma||\nabla_\gamma \log \pi_\gamma(\gamma)|} + \frac{\gamma^T \nabla_\gamma \log|\nabla_\gamma h^{-1}(\gamma)|}{|\gamma||\nabla_\gamma \log \pi_\gamma(\gamma)|}$$

Clearly the $\limsup$ of the right hand term in the sum is 0. As in the proof of Theorem 2, the left hand term in this sum can be rewritten as the sum of the following two terms

$$2\frac{h^{-1}(\gamma)^T\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)}{|\gamma||\nabla_\beta \log \pi_\gamma(\gamma)|} \tag{34}$$

$$2\frac{\big(\nabla_\gamma h^{-1}(\gamma)\gamma/2 - h^{-1}(\gamma)\big)^T\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)}{|\gamma||\nabla_\beta \log \pi_\gamma(\gamma)|} \tag{35}$$

Clearly the $\limsup$ of (35) is 0, since as $|\gamma|$ goes to ∞ the numerator is bounded, and the denominator is unbounded.

We can rewrite (34) as the product of the following three terms

$$2\frac{|h^{-1}(\gamma)|}{|\gamma||\nabla_\gamma h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)|} \tag{36}$$

$$\frac{h^{-1}(\gamma)^T\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)}{|h^{-1}(\gamma)|} \tag{37}$$

$$\frac{|\nabla h^{-1}(\gamma)\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)|}{|\nabla_\gamma \log \pi_\gamma(\gamma)|} \tag{38}$$

Since π_β has exponentially light tails, there exists a $\varepsilon > 0$ such that for $|\gamma|$ large enough (37) is bounded above by $-\varepsilon$. By Lemma 2 (38) goes to 1 as $|\gamma|$ goes to ∞. So if (36) is bounded away from 0 as $|\gamma|$ goes to ∞, the $\limsup$ of (34) is less than 0.

Now, the square of the denominator in (36) can be written as

$$\left[\sum_j \gamma_j^2\right]\left[\sum_j g'(\gamma_j)^2\Big(\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)\Big)_j^2\right]$$

which is bounded above by

$$m^2\left[\max_j \gamma_j^2\right]\left[\max_j g'(\gamma_j)^2\right]\left[\max_j\Big(\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)\Big)_j^2\right]$$

and since $|\big(\nabla_\beta \log \pi_\beta\big)h^{-1}(\gamma)|$ is bounded as $|\gamma|$ goes to ∞, the rightmost max is bounded above by some $b_1 > 0$ so the denominator of (36) is bounded above by

$$m\sqrt{b_1}\left[\max_j |\gamma_j|\right]\left[\max_j |g'(\gamma_j)|\right] \tag{39}$$

Likewise, the numerator of (36) below by

$$\max_j |g(\gamma_j)|. \tag{40}$$

As $|\gamma|$ goes to ∞ we have

$$j_0 = \arg\max_j |\gamma_j| = \arg\max_j |g'(\gamma_j)| = \arg\max_j |g(\gamma_j)|$$

and $|\gamma_{j_0}| > 1$. Therefore (36) is bounded below by

$$\frac{2}{m\sqrt{b_1}}\frac{\gamma_{j_0}^2 + 1/3}{2\gamma_{j_0}^2}$$

which is bounded below by $(m\sqrt{b_1})^{-1} > 0$. Therefore the $\limsup$ of (34) is less than 0, so (5) holds for π_γ. □

3.4 Geometric Ergodicity of a Random-Walk for the Transformed Density

Theorem 4. *Let π_β be an always positive density with continuous first derivatives that has exponentially light tails but is not a super-exponential density. Then if π_β satisfies conditions* (31) *and* (32) *and $\gamma = h(\beta)$ for h^{-1} defined in* (12) *then a random walk for π_γ with proposal density q satisfying* (4) *is geometrically ergodic.*

Proof. The proof follows directly from Theorems 1, 2, and 3. □

4 Application to Multinomial Logit Regression

4.1 The Model

Suppose that

$$Y \sim \text{Multinomial}(n, p)$$

where n is a positive integer and $p = (p_1, \ldots, p_k)^T$ satisfies the constraints

$$\sum_{i=1}^{k} p_1 = 1 \text{ and } \forall\, i,\ p_i \geq 0.$$

n gives the number of trials, and p_i gives the probability that a trial will result in category i. The pmf for $Y|n, p$ is

$$P(Y = y \mid n, p) = \frac{n!}{y_1! \cdots y_k!} p_1^{y_1} \cdots p_k^{y_k}. \tag{41}$$

Equation 41 gives rise to the canonical parametrization $\theta = (\theta_1, \ldots, \theta_k)^T$ where

$$\theta_j = \log(p_j) + c.$$

With the relation $e^c e^{\theta_j} = p_j$ we can rewrite (41) in canonical form

$$P(Y = y \mid n, p) = \frac{n!}{\prod_j y_j!} e^{\langle y, \theta \rangle - nC(\theta)} \tag{42}$$

where

$$C(\theta) = \log\Big(\sum_i e^{\theta_i}\Big).$$

θ is not directly of interest, instead we are interested in inference about the regression parameter, β. We replace θ with

$$M\beta.$$

M is the known model matrix. It has k rows, one for each outcome category. The number of columns of M depends on the number of predictor variables. For example, you may want to include all the one way effects (one for each level of each factor) and all the two way effects (one for each combination of levels from two factors). We will use m to refer to the number of effects included. Hence M is a $k \times m$ matrix and β is a vector of length m. In this report we will use $M_{j\cdot}$ to refer to the j^{th} row of M. Hence we have the relationship

$$\theta_j = M_{j\cdot}\beta.$$

Using this relationship we can rewrite (42) as

$$f(y|\beta) = \frac{n!}{\prod_j y_j!} e^{\langle y, M\beta \rangle - n \log\left(\sum_j e^{M_{j\cdot}\beta}\right)}. \tag{43}$$

We are working with a conjugate prior that adds made up data or "prior belief" to the observed data. If we use ξ to represent the cell counts and ν to represent the prior sample size, the posterior takes the form

$$\pi_\beta(\beta|y,n,\xi,\nu) \propto \exp\{ \ \langle y+\xi, M\beta\rangle - (n+\nu)\log\big(\sum_j e^{M_{j\cdot}\beta}\big) \ \}.$$

This is subject to the constraints $\sum_i \xi_i = \nu > 0$ and $\xi_i \geq 0$. As long as these constraints are satisfied there is no need for ξ or ν to take on integer values, although there is no harm done if they do. It may be convenient to define a probability distribution using η, where $\sum \eta_i = 1$ and $\eta_i \geq 0$ and replace ξ_i with $\eta_i \nu$. This will have no impact on the sampler, whichever form is more convenient should be used.

These proofs do not separate y from ξ or n from ν, so for simplicity we leave the posterior density as

$$\pi_\beta(\beta|y,n) \propto \exp\{ \ \langle y, M\beta\rangle - n\log\big(\sum_j e^{M_{j\cdot}\beta}\big) \ \} \tag{44}$$

and the gradient of the log posterior density as

$$\nabla_\beta \log \pi_\beta(\beta|y,n) = M^T y - n\sum_j M_{j\cdot}^T \frac{e^{M_{j\cdot}\beta}}{\sum_l e^{M_{l\cdot}\beta}}. \tag{45}$$

All of the proofs will still hold with y replaced by $y+\xi$ and n replaced by $n+\nu$ as long as the constraints on ξ and ν are satisfied.

4.2 Lemmas about π_β

Before we can prove our results about π_β as defined in (44) we need some facts and lemmas about π_β. To improve readability we will now use $\pi_\beta(\beta)$ instead of $\pi_\beta(\beta|y,n)$ to represent the posterior density of β. Likewise, we will use $\nabla_\beta \log \pi_\beta(\beta)$ instead of $\nabla_\beta \log \pi_\beta(\beta|y,n)$ to represent the gradient of the log posterior density of β.

4.2.1 Higher-dimensional derivative tests for convexity

Theorem 2.14 (higher-dimensional derivative tests) from Rockafeller and Wets (1998) states

Theorem 5. *For a differentiable function f on an open convex set $O \subset \mathbb{R}^n$, each of the following conditions is both necessary and sufficient for f to be convex on O:*

(a) $\langle x_1 - x_0, \nabla f(x_1) - \nabla f(x_0)\rangle \geq 0$ for all x_0 and x_1 in O;

(b) $f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle$ for all x and y in O;

(c) $\nabla^2 f(x)$ is positive-semidefinite for all x in O (f twice differentiable).

For strict convexity, a necessary and sufficient condition is (a) holding with strict inequality when $x_0 \neq x_1$, or (b) holding with strict inequality when $x \neq y$. A condition that is sufficient for strict convexity (but not necessary) is the positive definiteness of the Hessian matrix in (c) for all x in O.

4.2.2 Properties of π_β

We will use $\tilde{\beta}$ to denote the posterior mode of $\pi_\beta(\beta)$. These theorems depend on the fact that π_β is log concave (Geyer, 1990, Sec. 2.2) and that $\tilde{\beta}$ is the *unique* posterior mode of π_β.

For $\tilde{\beta}$ to be the unique posterior mode it is necessary for π_β to not have a direction of constancy. That is, there is no δ such that for all $s \in \mathbb{R}$ $\pi_\beta(\beta) = \pi_\beta(\beta + s\delta)$.

For π_β of the form (44) not having a direction of constancy means that $0 < y_i < n$ for each i and the column space of M does not include the vector with all components equal to 1. The condition $0 < y_i < n$ for each i simply means that there is 'data in all of the cells'.

Let S_1 denote the unit sphere in $\mathbb{R}^m$, that is

$$S_1 = \{ \delta \in \mathbb{R}^m \mid \delta^T \delta = 1 \} \tag{46}$$

Lemma 3. *If $s > 1$ and $\delta \in S_1$*

$$\langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + s\delta) \rangle \le \langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle \le 0$$

Proof. Since $\log \pi_\beta$ is concave, $-\log \pi_\beta$ is convex. We can apply Theorem 5 part (a) to $-\log \pi_\beta$ with $x_1 = \tilde{\beta} + s\delta$ and $x_0 = \tilde{\beta} + \delta$. This gives us

$$\begin{aligned}
0 &\le \langle \tilde{\beta} + s\delta - (\tilde{\beta} + \delta), -\nabla_\beta \log \pi_\beta(\tilde{\beta} + s\delta) - -\nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle \\
&= \langle (s-1)\delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) - \nabla_\beta \log \pi_\beta(\tilde{\beta} + s\delta) \rangle \\
&\Leftrightarrow (s-1)\langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + s\delta) \rangle \le (s-1)\langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle \\
&\Leftrightarrow \langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + s\delta) \rangle \le \langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle.
\end{aligned}$$

□

Lemma 4. *If $\delta \in S_1$ then there exists a $\varepsilon > 0$ such that*

$$\langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle \le -\varepsilon$$

Proof. Let $\delta \in S_1$ and apply Theorem 5 (b) with $y = \tilde{\beta}$, $x = \tilde{\beta} + \delta$ and $f = -\log \pi_\beta$ thus

$$\begin{aligned}
&\langle -\nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta), \tilde{\beta} - (\tilde{\beta} + \delta) \rangle - \log \pi_\beta(\tilde{\beta} + \delta) \le -\log \pi_\beta(\tilde{\beta}) \\
&\qquad \Leftrightarrow \langle \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta), \delta \rangle \le \log \pi_\beta(\tilde{\beta} + \delta) - \log \pi_\beta(\tilde{\beta}).
\end{aligned} \tag{47}$$

Since S_1 is compact, and $\log \pi_\beta$ is a continuous function, there exists a $\delta' \in S_1$ such that $\log \pi_\beta(\tilde{\beta} + \delta') = \sup_{\delta \in S_1} \log \pi_\beta(\tilde{\beta})$. Since $\tilde{\beta}$ is the unique posterior mode, and π_β does not have a direction of constancy $\log \pi_\beta(\tilde{\beta} + \delta') < \log \pi_\beta(\tilde{\beta})$ and $\varepsilon = \log \pi_\beta(\tilde{\beta}) - \log \pi_\beta(\tilde{\beta} + \delta') > 0$. Hence $\log \pi_\beta(\tilde{\beta} + \delta) - \log \pi_\beta(\tilde{\beta}) \le \log \pi_\beta(\tilde{\beta} + \delta') - \log \pi_\beta(\tilde{\beta}) = -\varepsilon < 0$. Combining this with (47) we get

$$\langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta} + \delta) \rangle \le -\varepsilon.$$

□

Now we are ready to prove the necessary conditions on π_β to use Theorem 4

Theorem 6. *Let π_β be a density of the form given in (44). If $\tilde{\beta}$ is the unique posterior mode of π_β and π_β does not have any directions of constancy then*

$$\limsup_{|\beta|\to\infty} \frac{\beta^T}{|\beta|} \nabla_\beta \log \pi_\beta(\beta) < 0 \tag{48}$$

Proof. For all β with $|\beta| > |\tilde{\beta}| + 1$, we can find a $s > 1$ and $\delta \in S_1$ such that $\beta = \tilde{\beta} + s\delta$. Thus instead of considering points β, we are considering points on the rays coming out from $\tilde{\beta}$. These rays combine to form $\mathbb{R}^m$, and we are only not considering β in the region where $|\beta| \le |\tilde{\beta}| + 1$. Hence as $|\beta|$ goes to ∞, it is always on some ray $\tilde{\beta} + s\delta$, with s going to ∞. Therefore showing that

$$\sup_{\delta\in S_1}\Big[\limsup_{s\to\infty} \frac{(\tilde{\beta}+s\delta)^T}{|\tilde{\beta}+s\delta|} \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta)\Big] < 0 \tag{49}$$

holds will show that (48) holds. Our strategy is to bound the inner lim sup in (49) above by a negative constant that does not depend on δ.

Splitting up the sum $(\tilde{\beta}+s\delta)^T$ and recalling that $\limsup(f(x)+g(x)) \le \limsup f(x) + \limsup g(x)$ we can bound the lim sup in (49) above by

$$\limsup_{s\to\infty} \frac{\tilde{\beta}^T}{|\tilde{\beta}+s\delta|} \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta) + \limsup_{s\to\infty} \frac{s\delta^T}{|\tilde{\beta}+s\delta|} \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta) \tag{50}$$

Recalling the definition of $\nabla_\beta \log \pi_\beta$ from (45) we have

$$\tilde{\beta}^T \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta) = \tilde{\beta}^T M^T y - n \sum_i \tilde{\beta}^T M_{i\cdot}^T \frac{e^{M_{i\cdot}(\tilde{\beta}+s\delta)}}{\sum_i e^{M_{i\cdot}(\tilde{\beta}+s\delta)}}$$

Clearly for all values of s and $\delta \in S_1$, $0 < \frac{e^{M_{i\cdot}(\tilde{\beta}+s\delta)}}{\sum_i e^{M_{i\cdot}(\tilde{\beta}+s\delta)}} < 1$. So $\tilde{\beta}^T \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta)$ is bounded above by $|\tilde{\beta} M^T y| + n \sum_i |\tilde{\beta}^T M_{i\cdot}^T|$. It follows that the first lim sup in (50) is bounded above by

$$\limsup_{s\to\infty} \frac{|\tilde{\beta}^T M^T y| + n\sum_i |\tilde{\beta}^T M_{i\cdot}^T|}{|\tilde{\beta}+s\delta|} = 0.$$

Now let $s > 1$ be given. We can rewrite the second lim sup in (50) as

$$\limsup_{s\to\infty} \frac{s}{|\tilde{\beta}+s\delta|} \langle \delta, \nabla_\beta \log \pi_\beta(\tilde{\beta}+s\delta)\rangle$$

and apply Lemma 3 and Lemma 4 to show that for some $\varepsilon > 0$ the second lim sup in (50) is bounded above by

$$\limsup_{s\to\infty} \frac{s}{|\tilde{\beta}+s\delta|}(-\varepsilon) = -\varepsilon < 0.$$

These bounds show that the lim sup in (49) is bounded above by $-\varepsilon$, which does not depend on δ, so the sup in (49) is also bounded above by $-\varepsilon$. Therefore the lim sup in (48) is bounded above by $-\varepsilon < 0$. □

Next we need to demonstrate that even though π_β has exponentially light tails, it is not a super-exponential density.

Theorem 7. *If π_β has the form given in* (44)*, even if $\tilde{\beta}$ is the unique posterior mode of π_β and π_β does not have any directions of constancy then π_β is not a super-exponential density.*

Proof. To prove that π_β is not a super-exponential density, it suffices to show that the lim sup in (48) is bounded below by some real number.

Let $\beta \in \mathbb{R}^m$ be given with $|\beta| > 0$. If we set $s = |\beta|$ and $\delta = \beta/|\beta|$ then $\beta = s\delta$ for $s > 0$ and $\delta \in S_1$. Recalling the definition of $\nabla_\beta \log \pi_\beta$ from (45) we can rewrite $\delta^T \nabla_\beta \log \pi_\beta(s\delta)$,

$$\delta^T \nabla_\beta \log \pi_\beta(s\delta) = \delta^T M^T y - n \sum_i \delta^T M_{i\cdot}^T \frac{e^{M_{i\cdot} s\delta}}{\sum_i e^{M_{i\cdot} s\delta}}.$$

Therefore for all s and δ $0 < \frac{e^{M_{i\cdot} s\delta}}{\sum_i e^{M_{i\cdot} s\delta}} < 1$, and $\delta^T \delta = 1$, the lim sup in (48) is bounded below by

$$-\sum_{i,j} |M_{ij} y_j| - n \sum_{i,j} |M_{ij}| > -\infty$$

so π_β is not a super-exponential density. □

Now we will demonstrate that π_β satisfies the requirements of Theorem 4.

Theorem 8. *If π_β is the density given in* (44)*, $0 < y_i$ for all i, and the column space of M does not contain the vector consisting of all 1s, then π_β satisfies the conditions for Theorem 4.*

Proof. Theorems 6 and 7 show that π_β has exponentially light tails but is not a super-exponential density. The fact that π_β is always positive and has continuous first derivatives follows directly from the form of π_β given in (44) and the form of $\nabla_\beta \log \pi_\beta$ given in (45). The fact that $|\nabla_\beta \log \pi_\beta(\beta)|$ is bounded away from 0 as $|\beta|$ goes to ∞ follows from Theorem 6. The fact that $|\nabla_\beta \log \pi_\beta(\beta)|$ is bounded above as $|\beta|$ goes to ∞ follows from Theorem 7. □

Since π_β as given in (44) satisfies the conditions of Theorem 4, we know that a random-walk Metropolis algorithm for π_β may or may not be geometrically ergodic, but a random-walk Metropolis algorithm with proposal density satisfying (4) for the density π_γ for $\gamma = h(\beta)$ for h^{-1} as defined in (6) will be geometrically ergodic. Thus we can draw inference the distribution of β using MCMC on the transformed density π_γ.

References

Geyer, C. J. (1990). Likelihood and exponential families. Unpublished Ph. D. Thesis, University of Washington. `http://purl.umn.edu/56330`.

Jarner, F. S. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications.* **85**, 341–361.

Mengersen, K. L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*. **24**, 101–121.

Roberts, G. O. and Tweedie R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*. **83**, 95–110.

Rockafeller, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*. Berlin Heidelberg: Springer-Verlag.