

VARIABLE TRANSFORMATION TO OBTAIN GEOMETRIC ERGODICITY IN THE RANDOM-WALK METROPOLIS ALGORITHM

BY LEIF T. JOHNSON AND CHARLES J. GEYER

Google Inc. and University of Minnesota

A random-walk Metropolis sampler is geometrically ergodic if its equilibrium density is super-exponentially light and satisfies a curvature condition [*Stochastic Process. Appl.* **85** (2000) 341–361]. Many applications, including Bayesian analysis with conjugate priors of logistic and Poisson regression and of log-linear models for categorical data result in posterior distributions that are not super-exponentially light. We show how to apply the change-of-variable formula for diffeomorphisms to obtain new densities that do satisfy the conditions for geometric ergodicity. Sampling the new variable and mapping the results back to the old gives a geometrically ergodic sampler for the original variable. This method of obtaining geometric ergodicity has very wide applicability.

1. Introduction. Markov chain Monte Carlo (MCMC) using the Metropolis–Hastings–Green algorithm [Metropolis et al. (1953), Hastings (1970), Green (1995)] or its special case the Gibbs sampler [Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990)] has become very widely used [Gilks, Richardson and Spiegelhalter (1996), Brooks et al. (2011)], especially after Gelfand and Smith (1990) pointed out that most Bayesian inference can be done using MCMC, and little can be done without it.

In ordinary, independent and identically distributed Monte Carlo (OMC), the asymptotic variance of estimates is easily calculated [Geyer (2011), Section 1.7]. In MCMC, the properties of estimates are more difficult to handle theoretically [Geyer (2011), Section 1.8]. A Markov chain central limit theorem (CLT) may or may not hold [Tierney (1994), Chan and Geyer (1994)]. If it does hold, the asymptotic variance of MCMC estimates is more difficult to estimate than for OMC estimates, but estimating the asymptotic variance of the MCMC estimates is doable [Geyer (1992), Flegal and Jones (2010), Geyer (2011), Section 1.10]. The CLT holds for all $L^{2+\varepsilon}$ functionals of a Markov chain if the Markov chain is geometrically ergodic [Chan and Geyer (1994)]. For a reversible Markov chain [Geyer (2011), Section 1.5] the CLT holds for all L^2 functionals if and only if the Markov chain is geometrically ergodic [Roberts and Rosenthal (1997)]. The CLT may hold

Received January 2012; revised June 2012.

MSC2010 subject classifications. Primary 60J05, 65C05; secondary 60J22.

Key words and phrases. Markov chain Monte Carlo, change of variable, exponential family, conjugate prior, Markov chain isomorphism, drift condition, Metropolis–Hastings–Green algorithm.

for some functionals of a Markov chain when the Markov chain is not geometrically ergodic [Gordin and Lifšic (1978), Maigret (1978), Kipnis and Varadhan (1986), Chan (1993), Tierney (1994), Chan and Geyer (1994), Roberts and Rosenthal (1997, 2004), Jones (2004)], but then it is usually very difficult to verify that a CLT exists for a given functional of the Markov chain. Thus geometric ergodicity is a very desirable property for a Markov chain to have. This is especially true because most instances of the Metropolis–Hastings–Green algorithm are reversible or can be made to be reversible [Geyer (2011), Sections 1.5, 1.12 and 1.17], so, as stated above, geometric ergodicity implies the CLT holds for all L^2 functionals of the Markov chain, which makes reversible geometrically ergodic MCMC just as good as OMC in this respect.

Geometric ergodicity also plays a key role in the theory of calculable nonasymptotic bounds for Markov chain estimators [Rosenthal (1995b), Łatuszyński and Niemiro (2011), Łatuszyński, Miasojedow and Niemiro (2012)], but is only half of what must be done to establish this type of result. The other half is establishing a minorization condition. The proof techniques involved in establishing geometric ergodicity and in establishing minorization conditions, however, have little in common. We deal only with establishing geometric ergodicity.

1.1. The random-walk Metropolis algorithm. The Metropolis–Hastings–Green algorithm generates a Markov chain having a specified invariant probability distribution. We restrict our attention to distributions of continuous random vectors, those having a density π with respect to Lebesgue measure on \mathbb{R}^k . If π is only known up to a normalizing constant, then the Metropolis–Hastings–Green algorithm still works.

We describe only the random-walk Metropolis algorithm [terminology introduced by Tierney (1994)]. This simulates a Markov chain X_1, X_2, \dots having π as an invariant distribution. It is determined by π and another function $q: \mathbb{R}^k \rightarrow \mathbb{R}$ that is a properly normalized probability density with respect to Lebesgue measure on \mathbb{R}^k and is symmetric about zero. Each iteration does the following three steps, where X_n is the state of the Markov chain before the iteration and X_{n+1} is the state after the iteration. Simulate Z_n having the distribution q , and set $Y_n = X_n + Z_n$. Calculate

$$(1) \quad a(X_n, Y_n) = \min(1, \pi(Y_n)/\pi(X_n)).$$

Set $X_{n+1} = Y_n$ with probability $a(X_n, Y_n)$, and set $X_{n+1} = X_n$ with probability $1 - a(X_n, Y_n)$.

The only requirement is $\pi(X_1) > 0$. The operation of the algorithm itself then ensures that $\pi(X_n) > 0$ almost surely for all n , so (1) always makes sense.

The proposal density q and target density π are arbitrary. The algorithm always produces a (not necessarily ergodic) reversible Markov chain having invariant density π regardless of what q is chosen. If q is everywhere positive, then

the Markov chain is necessarily ergodic [irreducible and positive Harris recurrent, Tierney (1994), Corollary 2].

The R package `mcmc` [Geyer and Johnson (2012)] provides a user-friendly implementation of the random-walk Metropolis algorithm combined with the variable transformation methodology described in this article in its `morph.metrop` function. The user provides an R function that evaluates $\log \pi$, and the `metrop` function in that package does the simulation. If the user correctly codes the function that evaluates $\log \pi$, then the `morph.metrop` function is guaranteed to simulate a reversible ergodic Markov chain having invariant density π . This gives an algorithm having an enormous range of application, which includes all Bayesian inference for models with continuous parameters and continuous prior distributions. No other computer package known to us combines this range of application with the correctness guarantees of the `mcmc` package, which are as strong as can be made about arbitrary user-specified target distributions.

1.2. Geometric ergodicity and random-walk Metropolis. A random-walk Metropolis sampler is not necessarily geometrically ergodic, but its geometric ergodicity has received more attention [Mengersen and Tweedie (1996), Roberts and Tweedie (1996), Jarner and Hansen (2000)] than any other MCMC sampler, except perhaps independence Metropolis–Hastings samplers, also terminology introduced by Tierney (1994), which are also studied in Mengersen and Tweedie (1996) and Roberts and Tweedie (1996). Independence Metropolis–Hastings samplers, however, do not have good properties, being either uniformly ergodic or not geometrically ergodic and uniformly ergodic only when its proposal distribution is particularly adapted to π in a way that is difficult to achieve (whenever independence samplers work, importance sampling also works, so MCMC is unnecessary).

To simplify the theory, Mengersen and Tweedie (1996), Roberts and Tweedie (1996) and Jarner and Hansen (2000) restrict attention to π that are strictly positive and continuously differentiable. In order to build on their results, we also adopt this restriction. The geometric ergodicity properties of the random-walk Metropolis algorithm are related to

$$(2) \quad \limsup_{|x| \rightarrow \infty} \frac{x}{|x|} \cdot \nabla \log \pi(x),$$

where the dot indicates inner product, and $|\cdot|$ denotes the Euclidean norm. We say π is *super-exponentially light* if (2) is $-\infty$, is *exponentially light* if (2) is negative and *sub-exponentially light* if (2) is zero.

None of these conditions are necessary for geometric ergodicity. A necessary condition for the geometric ergodicity of a random-walk Metropolis algorithm is that the target density π have a moment generating function [Jarner and Tweedie

(2003)]. It is possible for a density to have a moment generating function but not be even sub-exponentially light, for example, the unnormalized density

$$\pi(x) = e^{-|x|}(1 + \cos(x)), \quad x \in \mathbb{R}.$$

Following [Roberts and Tweedie \(1996\)](#) and [Järner and Hansen \(2000\)](#), we also restrict attention to q that are bounded away from zero in a neighborhood of zero. This includes the normal proposal distributions used by the R package `mcmc`.

THEOREM 1 [[Järner and Hansen \(2000\)](#), Theorem 4.3]. *Suppose π is a super-exponentially light density on \mathbb{R}^k that also satisfies*

$$(3) \quad \limsup_{|x| \rightarrow \infty} \frac{x}{|x|} \cdot \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0,$$

where the dot denotes inner product; then the random-walk Metropolis algorithm with q bounded away from zero on a neighborhood of zero is geometrically ergodic.

We say π *satisfies the curvature condition* to mean (3) holds. This means the contours of π are approximately locally linear near infinity.

Theorem 1, although useful, covers neither exponentially light densities, which arise in Bayesian categorical data analysis with canonical parameters and conjugate priors (Section 3.1), nor sub-exponentially light densities, which arise in Bayesian analysis of Cauchy location models using flat improper priors on the location parameters (Section 3.4). [Roberts and Tweedie \(1996\)](#) do cover exponentially light densities, but their theorems are very difficult to apply [[Järner and Hansen \(2000\)](#) show that [Roberts and Tweedie \(1996\)](#) incorrectly applied their own theorem in one case].

The key idea of this paper is to use the change-of-variable theorem in conjunction with Theorem 1 to get results that Theorem 1 does not give directly. Suppose π_β is the (possibly multivariate) target density of interest. We instead simulate a Markov chain having invariant density

$$(4) \quad \pi_\gamma(\gamma) = \pi_\beta(h(\gamma)) |\det \nabla h(\gamma)|,$$

where h is a diffeomorphism. If π_β is the density of the random vector β , then π_γ is the density of the random vector $\gamma = h^{-1}(\beta)$. We find conditions on the transformation h that make π_γ super-exponentially light and satisfy the curvature condition. Then by Theorem 1, the simulated Markov chain $\gamma_1, \gamma_2, \dots$ is geometrically ergodic. It is easy to see (Appendix A) that the Markov chain $\beta_i = h(\gamma_i)$, $i = 1, 2, \dots$, is also geometrically ergodic. Thus we achieve geometric ergodicity indirectly, doing a change-of-variable yielding a density that by Theorem 1 has a geometrically ergodic random-walk Metropolis sampler, sampling that distribution, and then using the inverse change-of-variable to get back to the variable of interest.

This indirect procedure has no virtues other than that Metropolis random-walk samplers are well-understood and user-friendly and that we have Theorem 1 to build on. There is other literature using drift conditions to prove geometric ergodicity of Markov chain samplers [Geyer and Møller (1994), Rosenthal (1995a), Hobert and Geyer (1998), Jones and Hobert (2004), Roy and Hobert (2007), Tan and Hobert (2009), Johnson and Jones (2010)] but for Gibbs samplers or other samplers for specific statistical models, hence not having the wide applicability of random-walk Metropolis samplers. There is also other literature about using variable transformation to improve the convergence properties of Markov chain samplers [Roberts and Sahu (1997), Papaspiliopoulos, Roberts and Sköld (2007), Papaspiliopoulos and Roberts (2008)] but for Gibbs samplers not having the wide applicability of random-walk Metropolis samplers.

It is important to understand that the necessary condition mentioned above [Jarner and Tweedie (2003)] places a limit on what can be done without variable transformation. If π_β does not have a moment generating function (any Student t distribution, e.g.), then no random-walk Metropolis sampler for it can be geometrically ergodic (no matter what proposal distribution is used). Thus if we use a random-walk Metropolis sampler, then we must also use variable transformation to obtain geometric ergodicity.

We call a function $h : \mathbb{R}^k \rightarrow \mathbb{R}^k$ *isotropic* if it has the form

$$(5) \quad h(\gamma) = \begin{cases} f(|\gamma|) \frac{\gamma}{|\gamma|}, & \gamma \neq 0, \\ 0, & \gamma = 0 \end{cases}$$

for some function $f : (0, \infty) \rightarrow (0, \infty)$. To simplify the theory, we restrict attention to h that are isotropic diffeomorphisms, meaning h and h^{-1} are both continuously differentiable, having the further property that $\det(\nabla h)$ and $\det(\nabla h^{-1})$ are also continuously differentiable.

As with the restriction to π that are strictly positive and continuously differentiable used by Mengersen and Tweedie (1996), Roberts and Tweedie (1996) and Jarner and Hansen (2000), this restriction is arbitrary. It is not necessary to achieve geometric ergodicity; it merely simplifies proofs. However, the proofs are already very complicated even with these two restrictions. Although both these restrictions could be relaxed, that would make the proofs even more complicated. Since many applications can be fit into our framework, perhaps after a change-of-variable to yield π_β that is strictly positive and continuously differentiable, we choose to not complicate our proofs further.

Isotropic transformations (5) shrink toward or expand away from the origin of the state space. In practice, they should be combined with translations so they can shrink toward or expand away from arbitrary points. Since translations induce isomorphic Markov chains (Appendix A), they do not affect the geometric ergodicity properties of random-walk Metropolis samplers. Hence we ignore them until Section 4.

Our variable-transformation method is easily implemented using the R package `mcmc` [Geyer and Johnson (2012)] because that package simulates Markov chains having equilibrium density π specified by a user-written function, which can incorporate a variable transformation, and outputs an arbitrary functional of the Markov chain specified by another user-written function, which can incorporate the inverse transformation.

A referee pointed out that one can think of our transformation method differently: as describing a Metropolis–Hastings algorithm in the original parameterization. This seems to avoid variable transformation but does not, because its proposals have the form $h(h^{-1}(\beta) + z)$, where β is the current state, and z is a simulation from the Metropolis q . This uses h and h^{-1} in every iteration, whereas the scheme we describe uses only h to run the Markov chain for γ and to map it back to β , needing h^{-1} only once to determine the initial state $\gamma_1 = h^{-1}(\beta_1)$ of the Markov chain. Nevertheless, it is of some theoretical interest that this provides hitherto unnoticed examples of geometrically ergodic Metropolis–Hastings algorithms.

2. Variable transformation.

2.1. Positivity and continuous differentiability. For the change-of-variable (4) we need to know when the transformed density π_γ is positive and continuously differentiable assuming the original density π_β has these properties. If h is a diffeomorphism, then the first term on the right-hand side will be continuously differentiable by the chain rule. Since ∇h^{-1} is the matrix inverse of ∇h by the inverse function theorem, $\det(\nabla h)$ can never be zero. Hence h being a diffeomorphism is enough to imply positivity of π_γ .

Since $\det(A)$ is continuous in A , being a polynomial function of the components of A , $\det(\nabla h)$ can never change sign. We restrict attention to h such that $\det(\nabla h)$ is always positive, so the absolute value in (4) is unnecessary. Then we have

$$(6) \quad \log \pi_\gamma(\gamma) = \log \pi_\beta(h(\gamma)) + \log \det(\nabla h(\gamma)),$$

$$(7) \quad \nabla \log \pi_\gamma(\gamma) = \nabla (\log \pi_\beta)(h(\gamma)) \nabla h(\gamma) + \nabla \log \det(\nabla h(\gamma)).$$

It is clear from (7) that $\log \pi_\gamma$, and hence π_γ is continuously differentiable if h is a diffeomorphism, and $\det(\nabla h)$ is continuously differentiable.

2.2. Isotropic functions. In the transformation method, the induced density, π_γ will need to satisfy the smoothness conditions of Theorem 1. We require the original density, π_β to satisfy the smoothness conditions of Theorem 1. The smoothness conditions will be satisfied for π_γ if the isotropic transformations are diffeomorphisms with continuously differentiable Jacobians. The assumptions of the following lemma provide conditions on isotropic functions to guarantee that π_γ is positive and continuously differentiable whenever π_β is.

LEMMA 1. Let $h: \mathbb{R}^k \rightarrow \mathbb{R}^k$ be an isotropic function given by (5) with $f: [0, \infty) \rightarrow [0, \infty)$ invertible and continuously differentiable with one-sided derivative at zero such that

$$(8) \quad f'(s) > 0, \quad s \geq 0.$$

Then

$$(9) \quad \frac{\gamma}{|\gamma|} = \frac{h(\gamma)}{|h(\gamma)|}, \quad \gamma \neq 0,$$

f is a diffeomorphism, h is a diffeomorphism and

$$(10) \quad h^{-1}(\beta) = \begin{cases} f^{-1}(|\beta|) \frac{\beta}{|\beta|}, & \beta \neq 0, \\ 0, & \beta = 0 \end{cases}$$

and

$$(11) \quad \nabla h(\gamma) = \frac{f(|\gamma|) \mathbf{I}_k}{|\gamma|} + \left[f'(|\gamma|) - \frac{f(|\gamma|)}{|\gamma|} \right] \frac{\gamma \gamma^T}{|\gamma|^2}, \quad \gamma \neq 0,$$

where \mathbf{I}_k is the $k \times k$ identity matrix, and

$$(12) \quad \nabla h(0) = f'(0) \mathbf{I}_k.$$

Moreover

$$(13) \quad \det(\nabla h(\gamma)) = \begin{cases} f'(|\gamma|) \left(\frac{f(|\gamma|)}{|\gamma|} \right)^{k-1}, & \gamma \neq 0, \\ f'(0)^k, & \gamma = 0 \end{cases}$$

and, under the additional assumption that f is twice continuously differentiable with one-sided derivatives at zero and

$$(14) \quad f''(0) = 0,$$

(13) is continuously differentiable.

The proof of this lemma is in Appendix B.

2.3. *Inducing lighter tails.* Define $f: [0, \infty) \rightarrow [0, \infty)$ by

$$(15) \quad f(x) = \begin{cases} x, & x < R, \\ x + (x - R)^p, & x \geq R, \end{cases}$$

where $R \geq 0$ and $p > 2$. It is clear that (15) satisfies the assumptions of Lemma 1.

THEOREM 2. Let π_β be an exponentially light density on \mathbb{R}^k , and let h be defined by (5) and (15). Then π_γ defined by (4) is super-exponentially light.

Proof of Theorem 2 is in Appendix C.

Now define $f : [0, \infty) \rightarrow [0, \infty)$ by

$$(16) \quad f(x) = \begin{cases} e^{bx} - \frac{e}{3}, & x > \frac{1}{b}, \\ x^3 \frac{b^3 e}{6} + x \frac{be}{2}, & x \leq \frac{1}{b}, \end{cases}$$

where $b > 0$. It is clear that (16) satisfies the assumptions of Lemma 1.

THEOREM 3. *Let π_β be a sub-exponentially light density on \mathbb{R}^k , and suppose there exist $\alpha > k$ and $R < \infty$ such that*

$$(17) \quad \frac{\beta}{|\beta|} \cdot \nabla \log \pi_\beta(\beta) \leq -\frac{\alpha}{|\beta|}, \quad |\beta| > R.$$

Let h be defined by (5) and (16). Then π_γ defined by (4) is exponentially light.

Proof of Theorem 3 is in Appendix C.

Condition (17) is close to sharp. For example, if π_β looks like a multivariate t distribution

$$(18) \quad \pi_\beta(t) = [1 + (t - \mu)^T \Sigma^{-1} (t - \mu)]^{-(v+k)/2}$$

[compare with (27) in Section 3.3], then (17) holds with $\alpha = k + v$, and (18) is integrable if and only if $v > 0$.

Moreover, an exponential-type isotropic transformation like (16) is necessary to obtain a super-exponentially light π_γ when π_β is a multivariate t distribution. Direct calculation shows that no polynomial-type isotropic transformation like (15) does the job.

COROLLARY 1. *Let π_β satisfy the conditions of Theorem 3, and let h be defined as the composition of those used in Theorems 2 and 3; that is, if we denote the h used in Theorem 2 by h_1 and denote the h used in Theorem 3 by h_2 , then in this corollary we are using $h = h_2 \circ h_1$ and the change of variable is $\gamma = h_1^{-1}(h_2^{-1}(\beta))$. Then π_γ defined by (4) is super-exponentially light.*

PROOF. The proof follows directly from Theorems 2 and 3. \square

2.4. Curvature conditions. As seen in Jarner and Hansen (2000), Example 5.4, being super-exponentially light is *not* a sufficient condition for the geometric ergodicity of a random-walk Metropolis algorithm. Jarner and Hansen (2000) provide sufficient conditions for super-exponentially light densities. In this section, we provide sufficient conditions for sub-exponentially light and exponentially light densities, such that, using the transformations from Section 2.3 the induced super-exponential densities will satisfy the Jarner and Hansen (2000) sufficient conditions.

THEOREM 4. *Let π_β be an exponentially light density on \mathbb{R}^k , and suppose that π_β satisfies either of the following conditions:*

- (i) π_β satisfies the curvature condition (3), or
- (ii) $|\nabla \log \pi_\beta(\beta)|$ is bounded as $|\beta|$ goes to infinity.

Let h be defined by (5) and (15). Then π_γ defined by (4) satisfies the curvature condition (3).

Proof of Theorem 4 is in Appendix D.

For exponentially light π_β , condition (ii) implies condition (i). In practice, condition (ii) may be easier to check than condition (i) (as in Section 3.1).

THEOREM 5. *Let π_β be a sub-exponentially light density on \mathbb{R}^k , and suppose there exist $\alpha > k$ and $R < \infty$ such that*

$$(19) \quad |\nabla \log \pi_\beta(\beta)| \leq \frac{\alpha}{|\beta|}, \quad |\beta| > R.$$

Let h be defined by (5) and (16). Then π_γ defined by (4) satisfies condition (ii) of Theorem 4 with β replaced by γ .

Proof of Theorem 5 is in Appendix D.

Condition (19), like (17), is close to sharp. If π_β has the form (18), then (19) holds with $\alpha = k + v$, and (18) is integrable if and only if $v > 0$.

COROLLARY 2. *Let π_β satisfy the conditions of Theorems 3 and 5, and let h be defined as the composition of those used in Theorems 4 and 5, that is, if we denote the h used in Theorem 4 by h_1 and denote the h used in Theorem 5 by h_2 , then in this corollary we are using $h = h_2 \circ h_1$ and the change of variable is $\gamma = h_1^{-1}(h_2^{-1}(\beta))$. Then π_γ defined by (4) satisfies the curvature condition (3).*

PROOF. This follows directly from Theorems 5 and 4. \square

To verify that a variable transformation (5) produces geometric ergodicity, one uses Theorems 2 and 4 when the given target density π_β is exponentially light. To verify that a variable transformation (5) produces geometric ergodicity, one uses Corollaries 1 and 2 when the given target density π_β is sub-exponentially light. (When the given target density π_β is super-exponentially light one does not need variable transformation to obtain geometric ergodicity if π_β also satisfies the curvature condition.)

3. Examples.

3.1. *Exponential families and conjugate priors.* In this section we study Bayesian inference for exponential families using conjugate priors, in particular, the case where the natural statistic is bounded in some direction, and the natural parameter space is all of \mathbb{R}^k . Examples include logistic regression, Poisson regression with log link function and log-linear models in categorical data analysis. In this case, we find that the posterior density, when it exists, is exponentially light and satisfies the curvature condition. Hence variable transformation using (5) and (15) makes the random-walk Metropolis sampler geometrically ergodic.

An exponential family is a statistical model having log likelihood of the form

$$y \cdot \beta - c(\beta),$$

where the dot denotes inner product, y is a vector statistic, β is a vector parameter and the function c is called the cumulant function of the family. A statistic y and parameter β that give a log likelihood of this form are called *natural* or *canonical*. If y_1, \dots, y_n are independent and identically distributed observations from the family and \bar{y}_n their average, then the log likelihood for the sample of size n is

$$n\bar{y}_n \cdot \beta - nc(\beta).$$

The log unnormalized posterior when using conjugate priors is

$$(20) \quad w(\beta) = (n\bar{y}_n + \nu\eta) \cdot \beta - (n + \nu)c(\beta),$$

where ν is a scalar hyperparameter, and η is a vector hyperparameter [Diaconis and Ylvisaker (1979), Section 2]. When simulating the posterior using MCMC, the unnormalized density of the target distribution is $\pi(\beta) = e^{w(\beta)}$.

The *convex support* of an exponential family is the smallest closed convex set containing the natural statistic with probability one. (This does not depend on which distribution in the exponential family we use because they are all mutually absolutely continuous.) Theorem 1 in Diaconis and Ylvisaker (1979) says that the posterior exists; that is, $e^{w(\beta)}$ is integrable, where $w(\beta)$ is given by (20), if and only if $n + \nu > 0$ and $(n\bar{y}_n + \nu\eta)/(n + \nu)$ is an interior point of the convex support. (Of course, this always happens when using a proper prior, i.e., when $\nu > 0$ and η/ν is an interior point of the convex support.)

Theorem 9.13 in Barndorff-Nielsen (1978) says that this same condition holds if and only if the log unnormalized posterior (20) achieves its maximum at a unique point, the posterior mode, call it β_n . (Ostensibly, this theorem applies only to log likelihoods of exponential families not to log unnormalized posteriors with conjugate priors, but since the latter have the same algebraic form as the former, it actually does apply to the latter.)

From the properties of exponential families [Barndorff-Nielsen (1978), Theorem 8.1],

$$(21) \quad \nabla c(\beta) = E_\beta(Y).$$

It follows that

$$(22) \quad \nabla \log \pi(\beta) = \nabla w(\beta) = n\bar{y}_n + v\eta - (n+v)E_\beta(Y).$$

Suppose that the natural statistic is bounded in some direction, that is, there exists a nonzero vector δ and real number b such that $y \cdot \delta \leq b$ for all y in the convex support. It follows that $E_\beta(Y) \cdot \delta \leq b$. Then

$$\begin{aligned} \limsup_{|\beta| \rightarrow \infty} \frac{\beta}{|\beta|} \cdot \nabla \log \pi(\beta) &\geq \limsup_{s \rightarrow \infty} \frac{s\delta}{|s\delta|} \cdot [n\bar{y}_n + v\eta - (n+v)E_{s\delta}(Y)] \\ &\geq \frac{(n\bar{y}_n + v\eta) \cdot \delta - (n+v)b}{|\delta|}. \end{aligned}$$

Hence (2) is not $-\infty$ and the target distribution is not super-exponentially light.

When the convex support has nonempty interior, the cumulant function c is strictly convex [Barndorff-Nielsen (1978), Theorem 7.1]. Hence (20) is a strictly concave function. It follows from this that ∇c is a strictly multivariate monotone function, that is,

$$(23) \quad [\nabla c(\beta_1) - \nabla c(\beta_2)] \cdot (\beta_1 - \beta_2) > 0, \quad \beta_1 \neq \beta_2$$

[Rockafellar and Wets (1998), Theorem 2.14 and Chapter 12]. It follows that

$$(24) \quad \nabla w(\beta) \cdot \frac{\beta - \tilde{\beta}_n}{|\beta - \tilde{\beta}_n|} < 0, \quad \beta \neq \tilde{\beta}_n,$$

where w is given by (20), because $\nabla w(\tilde{\beta}_n) = 0$. Let B denote the boundary and E denote the exterior of the ball of unit radius centered at $\tilde{\beta}_n$. Since c is infinitely differentiable [Barndorff-Nielsen (1978), Theorem 7.2], so is w , and the left-hand side of (24) is a continuous function of β . Since B is compact, the left-hand side of (24) achieves its maximum over B , which must be negative, say $-\varepsilon$. For any $\beta \in E$ we have $t\beta + (1-t)\tilde{\beta}_n \in B$ when $t = 1/|\beta - \tilde{\beta}_n|$. By (23) we have

$$[\nabla w(\beta) - \nabla w(t\beta + (1-t)\tilde{\beta}_n)] \cdot \frac{\beta - \tilde{\beta}_n}{|\beta - \tilde{\beta}_n|} < 0$$

because

$$\beta - [t\beta + (1-t)\tilde{\beta}_n] = (1-t)(\beta - \tilde{\beta}_n)$$

is parallel to $\beta - \tilde{\beta}_n$. Thus

$$\nabla w(\beta) \cdot \frac{\beta - \tilde{\beta}_n}{|\beta - \tilde{\beta}_n|} < -\varepsilon, \quad \beta \in E$$

and

$$\limsup_{\beta \rightarrow \infty} \nabla w(\beta) \cdot \frac{\beta - \tilde{\beta}_n}{|\beta - \tilde{\beta}_n|} \leq -\varepsilon,$$

and this is easily seen to be equivalent to the unnormalized density (20) being exponentially light.

Now we check the curvature condition (3) for exponential families. In case the natural statistic is bounded in all directions, as in logistic regression and log-linear models, the curvature condition follows directly because the family satisfies condition (ii) of Theorem 4 because $\nabla \log \pi(\beta)$ is (22), and this is bounded. In case the natural statistic is bounded in some directions but not all directions, as in Poisson regression, we have to work harder and use condition (i) of Theorem 4. Because

$$\nabla \log \pi(\beta) = \frac{\nabla \pi(\beta)}{\pi(\beta)},$$

we have

$$\frac{\nabla \pi(\beta)}{|\nabla \pi(\beta)|} = \frac{\nabla w(\beta)}{|\nabla w(\beta)|},$$

where $\nabla w(\beta)$ is given by (22). And from (24) and $\nabla w(\beta) \neq 0$ for $\beta \neq \tilde{\beta}_n$, we obtain

$$(25) \quad \frac{\nabla w(\beta)}{|\nabla w(\beta)|} \cdot \frac{\beta - \tilde{\beta}_n}{|\beta - \tilde{\beta}_n|} < 0, \quad \beta \neq \tilde{\beta}_n,$$

and the rest of the proof that π satisfies the curvature condition is just like the proof that it is exponentially light given above except that (25) replaces (24).

3.2. Multinomial logit regresion with a conjugate prior. This example is a special case of the example in Section 3.1.

In multinomial logit regression, using a conjugate prior is equivalent to adding prior counts to the data cells. For observations $1, \dots, L$, represent these prior counts as $\xi_l v_l$ where ξ_l is a vector giving the prior probability for each response for the l th observation, and v_l is the prior sample size. For the l th observation, let the vector Y^l represent the counts in each response category, $N^l = \sum_i Y_i^l$ be the sample size and M^l be the model matrix. The log unnormalized posterior density for the regression parameter β is given by

$$(26) \quad \pi(\beta|y, n, \xi, v) \propto \exp \left\{ \sum_{l=1}^L (y^l + \xi^l v^l) \cdot M^l \beta - (n^l + v^l) \log \left(\sum_j e^{M_j^l \beta} \right) \right\},$$

where M_j^l is the j th row of the matrix M^l . So long as $y_i^l + \xi_i^l v^l$ is positive for all i and l —there is data (actual plus prior) in all cells— π will be exponentially light, and satisfy condition (3). Hence a random-walk Metropolis algorithm for the density induced by the approach in Theorems 2 and 4 will be geometrically ergodic.

3.3. *Multivariate T distributions.* The density of a multivariate t distribution on \mathbb{R}^k with v degrees of freedom, location parameter vector μ and scale parameter matrix Σ is given by

$$(27) \quad \pi_\beta(t) = \frac{\Gamma[(v+k)/2]}{\Gamma[v/2](v\pi)^{k/2} \det(\Sigma)^{1/2}} \left[1 + \frac{1}{v}(t - \mu)^T \Sigma^{-1}(t - \mu) \right]^{-(v+k)/2}$$

so

$$(28) \quad \nabla \log \pi_\beta(t) = \frac{-(v+k)\Sigma^{-1}(t - \mu)}{v + (t - \mu)^T \Sigma^{-1}(t - \mu)},$$

which implies

$$(29) \quad t \cdot \nabla \log \pi_\beta(t) \rightarrow -(v+k), \quad \text{as } t \rightarrow \infty,$$

so (27) is sub-exponentially light.

The condition of Theorem 3 is also implied by (29). To check the condition of Theorem 5 we calculate

$$|\nabla \log \pi_\beta(t)|^2 \leq \frac{(v+k)^2 \lambda_{\max}^2 |t - \mu|^2}{(\lambda_{\min} |t - \mu|^2)^2},$$

where λ_{\max} and λ_{\min} are the largest and smallest eigenvalues of Σ^{-1} . Hence

$$|\nabla \log \pi_\beta(t)| \leq \frac{(v+k)\lambda_{\max}}{\lambda_{\min} |t - \mu|},$$

and the condition of Theorem 5 also holds. So a random-walk Metropolis algorithm for the induced density π_γ that uses the transformation described in Corollaries 1 and 2 will be geometrically ergodic, and the inverse transformed Markov chain will be geometrically ergodic for π_β . Since the multivariate t distribution does not have a moment generating function, no random-walk Metropolis algorithm for π_β is geometrically ergodic [Jarner and Tweedie (2003)]. Variable transformation is essential.

The case $k = 1$ gives the univariate t distribution, which has been widely used as an example of a Harris ergodic random-walk Metropolis algorithm that is not geometrically ergodic [Mengersen and Tweedie (1996), Jarner and Hansen (2000), Jarner and Tweedie (2003), Jarner and Roberts (2007)].

3.4. *Cauchy location models and flat priors.* The t distribution with one degree of freedom is the Cauchy distribution. Consider a Cauchy location family with flat prior, so the posterior density for sample size one is again a Cauchy distribution

$$\pi_\beta(\mu) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \mu)^2},$$

and, this being a special case of the preceding section, this density is sub-exponentially light.

For a sample of size n the unnormalized posterior density is

$$\pi_\beta(\mu) = \prod_{i=1}^n \frac{1}{1 + (x_i - \mu)^2}$$

and the posterior distribution is no longer a brand name distribution. It is still easily shown to be sub-exponentially light and to satisfy the conditions of Theorems 3 and 5.

4. Discussion. The transformations in Theorems 2 and 3 will always induce a density with tails at least as light as the original density. If the original density satisfies the curvature condition, then the transformation using the transformation from Theorem 2 will induce a density that satisfies the curvature condition. Thus applying the transformation from Theorem 2 to a super-exponentially light density that satisfies the curvature condition will induce another super-exponentially light density that satisfies the curvature condition. We do not recommend transformation when the original density already satisfies the conditions of Theorem 1, but it seems this will do no harm.

The transformation method introduced here can be mixed blessing. It can produce geometric ergodicity, but may cause other problems. For example, π_γ given by (4) can be multimodal when π_β is unimodal. Thus we want a less extreme member of the family of transformations that does the job. The idea is to pull in the tails enough to get geometric ergodicity without much affecting the main part of the distribution. Although very extreme transformations work in theory, they are problematic in practice due to inexactness of computer arithmetic.

As mentioned in the Introduction, in practice one combines the transformations introduced in Section 2.3 with translations. Let t_λ denote the translation $x \mapsto x + \lambda$. Then in the exponentially light π_β case, we use the transformation $h = t_\lambda \circ h_{R,p}$, where $h_{R,p}$ is the h defined by (5) and (15), so the change-of-variable is $\gamma = h_{R,p}^{-1}(\beta - \lambda)$. This gives users three adjustable constants, λ , R and p , to experiment with to improve the mixing of the sampler. If π_β satisfies the assumptions of Theorems 2 and 4, then any valid values of λ , R and p result in a geometrically ergodic sampler. Observe that the restriction of this h to the ball of radius R centered at λ is a translation, which does not affect the shape of the distribution. Thus one wants to choose λ near the center of the distribution (perhaps the mode of π_β , if it has one) and R large enough so that a large part of the probability is in this ball where the shape is unchanged. The parameter p should always be chosen to be small, say 3 or 2.5 (recall $p > 2$ is required), 3 is a good choice as then f has a closed-form expression for its inverse.

In the sub-exponentially light π_β case, we use the transformation $h = t_\lambda \circ h_b \circ h_{R,p}$, where h_b is the h defined by (5) and (16), and the other two transformations are as above, so the change-of-variable is $\gamma = h_{R,p}^{-1}(h_b^{-1}(\beta - \lambda))$. This gives users four adjustable constants, λ , R , p and b to experiment with to improve the mixing

of the sampler. If π_β satisfies the assumptions of Corollaries 1 and 2, then any valid values of λ , R , p and b result in a geometrically ergodic sampler. One should choose the first three as discussed above, and b should be chosen to be small, say 0.1 or 0.01.

Admittedly, our methods do not guarantee geometric ergodicity without any theoretical analysis. Users must understand the tail behavior of the target distribution in order to select the correct transformation. For distributions with well behaved tails, this analysis may be easy, as in our examples. We can say that our methods are no more difficult to apply than the current state of the art [Jarner and Hansen (2000)] and are applicable to a much larger class of models.

APPENDIX A: ISOMORPHIC MARKOV CHAINS

We say measurable spaces are *isomorphic* if there is an invertible bimeasurable mapping between them (h bimeasurable means both h and h^{-1} are measurable). We say probability spaces (S, \mathcal{A}, P) and (T, \mathcal{B}, Q) are *isomorphic* if there is an invertible bimeasurable mapping $h: S \rightarrow T$ such that $P = Q \circ h$, meaning

$$P(A) = Q(h(A)), \quad A \in \mathcal{A},$$

which also implies $Q = P \circ h^{-1}$. We say Markov chains on state spaces (S, \mathcal{A}) and (T, \mathcal{B}) are *isomorphic* if there is an invertible bimeasurable mapping $h: S \rightarrow T$ such that the corresponding initial distributions μ and ν and the transition probability kernels P and Q satisfy $\mu = \nu \circ h$ and

$$(30) \quad P(x, A) = Q(h(x), h(A)), \quad x \in S \text{ and } A \in \mathcal{A}.$$

By the change-of-variable theorem for measures, (30) implies

$$(31) \quad P^n(x, A) = Q^n(h(x), h(A)), \quad n \in \mathbb{N} \text{ and } x \in S \text{ and } A \in \mathcal{A}.$$

It follows that P has an irreducibility measure if and only if Q has an irreducibility measure. It also follows from the change-of-variable theorem that η is an invariant measure for P if and only if $\eta \circ h^{-1}$ is an invariant measure for Q . Thus P is null recurrent if and only if Q is, and P is positive recurrent if and only if Q is. Also P is reversible with respect to η if and only if Q is reversible with respect to $\eta \circ h^{-1}$.

For Harris recurrence we use the criterion that a recurrent Markov chain is Harris if and only if every bounded harmonic function is constant [Nummelin (1984), Theorem 3.8 combined with his Proposition 3.9 and Theorem 8.0.1 of Meyn and Tweedie (2009)]. A function g is *harmonic* for a kernel P if $g = Pg$, meaning

$$g(x) = \int P(x, dy)g(y), \quad x \in S.$$

It is clear that g is harmonic for P if and only if $g \circ h^{-1}$ is harmonic for Q . Thus P is Harris recurrent if and only if Q is.

Suppose P is irreducible and periodic. This means [Meyn and Tweedie (2009), Proposition 5.4.1] there are disjoint sets D_0, \dots, D_{d-1} with $d \geq 2$ that are a partition of S such that

$$P(x, D_{i+1 \bmod d}) = 1, \quad x \in D_i, \quad i = 0, \dots, d-1.$$

But then

$$Q(y, h^{-1}(D_{i+1 \bmod d})) = 1, \quad y \in h^{-1}(D_i), \quad i = 0, \dots, d-1,$$

and the sets $h^{-1}(D_i)$ partition T , so Q is also periodic. Thus isomorphic irreducible Markov chains are both periodic or both aperiodic.

Finally suppose π is an invariant probability measure for P , and μ is any probability measure on the state space. Then $\psi = \pi \circ h^{-1}$ is an invariant probability measure for Q , and it is clear that

$$\|\pi - \mu P^n\| = \|\psi - \nu Q^n\|, \quad n \in \mathbb{N},$$

where $\|\cdot\|$ denotes total variation norm and $\nu = \mu \circ h^{-1}$. A Markov chain is *geometrically ergodic* if there exists a nonnegative-real-valued function M and constant $r < 1$ such that

$$(32) \quad \|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)r^n, \quad \text{for all } x$$

[Meyn and Tweedie (2009), Chapter 15]. If M is bounded, then the Markov chain is *uniformly ergodic* [Meyn and Tweedie (2009), Chapter 16]. If (32) holds with r^n replaced by n^r for some $r < 0$, then the Markov chain is *polynomially ergodic* [Jarner and Roberts (2002)]. Thus, if a Markov chain is polynomially ergodic, geometrically ergodic, or uniformly ergodic, then any isomorphic Markov chain has the same property.

The following summarizes the discussion in this appendix.

THEOREM 6 (Isomorphic Markov chains). *If a Markov chain has one of the following properties, irreducibility, reversibility, null recurrence, positive recurrence, Harris recurrence, aperiodicity, polynomial ergodicity, geometric ergodicity, uniform ergodicity, then so does any isomorphic Markov chain.*

APPENDIX B: PROOF OF LEMMA 1

That f is a diffeomorphism follows from the inverse function theorem

$$\frac{df^{-1}(t)}{dt} = \frac{1}{f'(s)} \quad \text{whenever } t = f(s)$$

and (8). It is clear from (5) that $|h(\gamma)| = f(|\gamma|)$ for all γ , from which (9), (10) and the invertibility of h follow.

Now for $\gamma \neq 0$ we have

$$\frac{\partial}{\partial \gamma_k} \left(\sum_{i=1}^d \gamma_i^2 \right)^{1/2} = \left(\sum_{i=1}^d \gamma_i^2 \right)^{-1/2} \gamma_k$$

so

$$\nabla |\gamma| = \frac{\gamma^T}{|\gamma|},$$

and now (11) follows straightforwardly from (5), and it is clear that h is continuously differentiable everywhere except perhaps at zero and similarly for h^{-1} .

The term in square brackets on the right-hand side of (11) goes to zero as $|\gamma| \rightarrow 0$ by the definition of derivative and that the term that multiplies it is bounded, thus, if we can show (12), then ∇h is also continuous at zero. By the definition of derivative, what must be shown to prove (12) is that

$$\begin{aligned} \frac{h(\gamma) - f'(0)\gamma}{|\gamma|} &= \frac{f(|\gamma|)(\gamma/|\gamma|) - f'(0)\gamma}{|\gamma|} \\ &= \left[\frac{f(|\gamma|)}{|\gamma|} - f'(0) \right] \frac{\gamma}{|\gamma|} \end{aligned}$$

converges to zero as $\gamma \rightarrow 0$. Since the term in square brackets converges to zero by the definition of derivative and $\gamma/|\gamma|$ is bounded, this proves (12). Since the formulas for h and h^{-1} have the same form, this shows h is a diffeomorphism.

The determinant of a symmetric matrix is the product of its eigenvalues [Harville (1997), Theorem 21.6.1]. First, γ is an eigenvector of $\nabla h(\gamma)$ with eigenvalue $f'(|\gamma|)$. Second, any vector v orthogonal to γ is also an eigenvector of $\nabla h(\gamma)$ with eigenvalue $f(|\gamma|)/|\gamma|$ when $\gamma \neq 0$ and eigenvalue $f'(0)$ when $\gamma = 0$. Since the subspace orthogonal to γ has dimension $k - 1$, the multiplicity of the second kind of eigenvalue is $k - 1$. This proves (13).

For $\gamma \neq 0$ we have

$$\begin{aligned} \nabla \det(\nabla h(\gamma)) &= f''(|\gamma|) \left(\frac{f(|\gamma|)}{|\gamma|} \right)^{k-1} \frac{\gamma^T}{|\gamma|} \\ (33) \quad &+ (k-1)f'(|\gamma|) \left(\frac{f(|\gamma|)}{|\gamma|} \right)^{k-2} \left[\frac{f'(|\gamma|)}{|\gamma|} - \frac{f(|\gamma|)}{|\gamma|^2} \right] \gamma^T. \end{aligned}$$

Since (13) depends on γ only through $|\gamma|$, it has circular contours, and we must have

$$(34) \quad \nabla \det(\nabla h(0)) = 0$$

if the derivative exists. We claim the derivative (34) does exist, and (13) is continuously differentiable under the “additional assumptions” about second derivatives

of f of the lemma. To prove this claim we need to first show that (33) converges to zero as $\gamma \rightarrow 0$ and second show that (34) is the derivative at zero.

Except for the behavior of the term in square brackets, the limit of (33) is obvious from $f(s)/s \rightarrow f'(0)$ as $s \rightarrow 0$ and $\gamma/|\gamma|$ being bounded. For the term in square brackets we use Taylor's theorem [Stromberg (1981), Theorem 4.34]

$$\begin{aligned} f(s) &= cs + o(s^2), \\ f'(s) &= c + o(s), \end{aligned}$$

where $c = f'(0)$, so

$$\frac{f'(s)}{s} - \frac{f(s)}{s^2} = o(1),$$

and the term in square brackets in (33) goes to zero as $\gamma \rightarrow 0$ proving that all of (33) goes to zero as $\gamma \rightarrow 0$.

What must be shown to establish (34) is that

$$\frac{\det(\nabla h(\gamma)) - \det(\nabla h(0))}{|\gamma|} = \frac{1}{|\gamma|} \left[f'(|\gamma|) \left[\frac{f(|\gamma|)}{|\gamma|} \right]^{k-1} - [f'(0)]^k \right]$$

converges to zero as $\gamma \rightarrow 0$. Applying L'Hospital's rule, we have

$$\begin{aligned} \lim_{s \downarrow 0} \frac{f'(s)[f(s)/s]^{k-1} - [f'(0)]^k}{s} \\ = \lim_{s \downarrow 0} \left[f''(s) \left[\frac{f(s)}{s} \right]^{k-1} + f'(s)(k-1) \left[\frac{f(s)}{s} \right]^{k-2} \left(\frac{f'(s)}{s} - \frac{f(s)}{s^2} \right) \right], \end{aligned}$$

and we have already shown that the limit on the right-hand side is zero.

APPENDIX C: PROOFS FROM SECTION 2.3

Before we prove Theorem 2 we need two additional lemmas.

LEMMA 2. *Let h be defined by (5) and (15). Then*

$$(35) \quad \lim_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \nabla \log \det(\nabla h(\gamma)) = 0,$$

where the dot indicates inner product.

PROOF. Recalling the value of $\det(\nabla h(\gamma))$ for $\gamma \neq 0$ from (13) we can rewrite the dot product in (35) as

$$(36) \quad \frac{f''(|\gamma|)}{f'(|\gamma|)} + (k-1) \left(\frac{f'(|\gamma|)}{f'(|\gamma|)} - \frac{1}{|\gamma|} \right).$$

From (15) for $|\gamma| > R$ we have

$$(37) \quad f'(x) = 1 + p(x - R)^{p-1},$$

$$(38) \quad f''(x) = p(p-1)(x - R)^{p-2}$$

and, plugging these into (36), we see that, because $p > 2$, all terms in (36) go to zero like $|\gamma|^{-1}$ as $|\gamma| \rightarrow \infty$. \square

LEMMA 3. *Under the assumptions of Lemma 1,*

$$(39) \quad \nabla h(\gamma)\gamma = f'(|\gamma|)\gamma, \quad \gamma \in \mathbb{R}^k,$$

$$(40) \quad [\nabla h(\gamma)]^2 = \frac{f(|\gamma|)^2}{|\gamma|^2} \mathbf{I}_k + \left[f'(|\gamma|)^2 - \frac{f(|\gamma|)^2}{|\gamma|^2} \right] \frac{\gamma\gamma^T}{|\gamma|^2}, \quad \gamma \neq 0,$$

$\nabla h(\gamma)$ being a symmetric matrix, and

$$(41) \quad x^T [\nabla h(\gamma)]^2 x = \frac{f(|\gamma|)^2}{|\gamma|^2} |x|^2 + \left[f'(|\gamma|)^2 - \frac{f(|\gamma|)^2}{|\gamma|^2} \right] \left(\frac{h(\gamma) \cdot x}{|h(\gamma)|} \right)^2, \\ x \in \mathbb{R}^k, \gamma \neq 0.$$

PROOF. From (11) and (12), we straightforwardly obtain (39) and for $\gamma \neq 0$

$$(42) \quad [\nabla h(\gamma)]^2 = \nabla h(\gamma) \left(\frac{f(|\gamma|)}{|\gamma|} \mathbf{I}_k + \left[\frac{f'(|\gamma|)}{|\gamma|^2} - \frac{f(|\gamma|)}{|\gamma|^3} \right] \gamma\gamma^T \right) \\ = \frac{f(|\gamma|)}{|\gamma|} \nabla h(\gamma) + \left[\frac{f'(|\gamma|)^2}{|\gamma|^2} - \frac{f(|\gamma|)f'(|\gamma|)}{|\gamma|^3} \right] \gamma\gamma^T$$

and

$$\frac{f(|\gamma|)}{|\gamma|} \nabla h(\gamma) = \frac{f(|\gamma|)^2}{|\gamma|^2} \mathbf{I}_k + \left[\frac{f'(|\gamma|)f(|\gamma|)}{|\gamma|^3} - \frac{f(|\gamma|)^2}{|\gamma|^4} \right] \gamma\gamma^T,$$

which plugged into (42) gives (40), and (41) is straightforward from (40). \square

PROOF OF THEOREM 2. Since $\nabla h(\gamma)$ is a symmetric matrix, it follows from (7) that

$$\gamma \cdot \nabla \log \pi_\gamma(\gamma) = \nabla h(\gamma)\gamma \cdot \log \pi_\beta(h(\gamma)) + \gamma \cdot \nabla \log \det(\nabla h(\gamma)).$$

Hence we can bound (2) by the sum of

$$(43) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\nabla h(\gamma)\gamma}{|\gamma|} \cdot \nabla \log \pi_\beta(h(\gamma))$$

and

$$(44) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \nabla \log \det(\nabla h(\gamma)).$$

It follows from (9) and (39) that for large $|\gamma|$ the dot product in (43) can be rewritten as

$$(45) \quad f'(|\gamma|) \frac{h(\gamma)}{|h(\gamma)|} \cdot \nabla \log \pi_\beta(h(\gamma)).$$

Since $f'(|\gamma|)$ is always positive, and π_β is exponentially light, there is an $\varepsilon > 0$ such that (45) is bounded above by $-f'_1(|\gamma|)\varepsilon$. It is clear that $f'(|\gamma|) \rightarrow \infty$ as $|\gamma| \rightarrow \infty$, so (43) is equal to $-\infty$. It follows from Lemma 2 that (44) is equal to zero, so (2) is equal to $-\infty$ and π_γ is a super-exponentially light density. \square

Before we prove Theorem 3 we need a lemma.

LEMMA 4. *Let h be defined by (5) and (16). Then*

$$(46) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \nabla \log \det(\nabla h(\gamma)) = bk,$$

where the dot indicates inner product.

PROOF. As in the proof of Lemma 2, the dot product in (46) can be written as (36). Clearly, $(k-1)/|\gamma|$ goes to zero as $|\gamma|$ goes to infinity. Hence, (46) is equal to

$$(47) \quad \limsup_{x \rightarrow \infty} \left[\frac{f''(x)}{f'(x)} + (k-1) \frac{f'(x)}{f(x)} \right]$$

if the limit exists. For $x > 1/b$, it follows from (16) that

$$\begin{aligned} f'(x) &= be^{bx}, \\ f''(x) &= b^2 e^{bx} \end{aligned}$$

and plugging these into (47) gives

$$\limsup_{x \rightarrow \infty} \left[\frac{b^2 e^{bx}}{be^{bx}} + (k-1) \frac{be^{bx}}{e^{bx} - e/3} \right],$$

which equals bk . \square

PROOF OF THEOREM 3. As in the proof of Theorem 2, (2) can be rewritten as the sum of (43) and (44), and for large $|\gamma|$ the dot product in (43) can be rewritten as (45). By (17) and the fact that $|h(\gamma)| = f(|\gamma|)$, (45) is bounded above

$$\limsup_{|\gamma| \rightarrow \infty} \left(-\alpha \frac{f'(|\gamma|)}{f(|\gamma|)} \right),$$

which when f is given by (16) is equal to $-b\alpha$. It follows that the limit superior in (2) is bounded above by $-b(\alpha - k)$. Since $\alpha > k$, this upper bound is less than 0, so π_γ is exponentially light. \square

APPENDIX D: PROOFS FROM SECTION 2.4

Some lemmas are needed to prove the curvature conditions for exponentially light densities.

LEMMA 5. *Let π_β be an exponentially light density on \mathbb{R}^k , and let h be defined by (5) and (15). Then*

$$(48) \quad |\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)| \rightarrow \infty \quad \text{as } |\gamma| \rightarrow \infty,$$

and π_γ defined by (4) has the property

$$(49) \quad \lim_{|\gamma| \rightarrow \infty} \frac{|\nabla \log \pi_\gamma(\gamma)|}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|} = 1.$$

PROOF. The square of the left-hand side of (48) is, by (41),

$$(50) \quad \frac{f(|\gamma|)^2}{|\gamma|^2} |\nabla \log \pi_\beta(h(\gamma))|^2 + \left[f'(|\gamma|)^2 - \frac{f(|\gamma|)^2}{|\gamma|^2} \right] \left(\frac{h(\gamma) \cdot \nabla \log \pi_\beta(h(\gamma))}{|h(\gamma)|} \right)^2,$$

hence (48) holds if and only if (50) goes to infinity. Since the left-hand term of (50) is nonnegative, it is sufficient to show that the right-hand term goes to infinity to show that all of (50) goes to infinity. By assumption π_β is exponentially light, and since $|h(\gamma)| = f(|\gamma|)$, there exists an $\varepsilon > 0$ and $M < \infty$ such that

$$\frac{h(\gamma) \cdot \nabla \log \pi_\beta(h(\gamma))}{|h(\gamma)|} \leq -\varepsilon, \quad |\gamma| \geq M.$$

Thus in order to prove (50) goes to infinity as $|\gamma|$ goes to infinity, it is sufficient to prove that the term in square brackets in (50) goes to infinity. Plugging in the definitions of f and f' from (15) and (37) for large x , we obtain

$$\begin{aligned} f'(x)^2 - \frac{f(x)^2}{x^2} &= [1 + p(x - R)^{p-1}]^2 - \frac{[x + (x - R)^p]^2}{x^2} \\ &= (p^2 - 1)x^{2p-2} + o(x^{2p-2}), \end{aligned}$$

and since $p > 2$ by assumption, this goes to infinity as x goes to infinity; hence (50) goes to infinity as $|\gamma|$ goes to infinity and (48) holds.

By (7), showing that (49) is true only requires showing that

$$(51) \quad \lim_{|\gamma| \rightarrow \infty} \frac{|\nabla \log \det(\nabla h(\gamma))|}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|} = 0.$$

It follows from (13) that for $\gamma \neq 0$,

$$\log \det(\nabla h(\gamma)) = \log f'(|\gamma|) + (k - 1) \log \left(\frac{f(|\gamma|)}{|\gamma|} \right)$$

and

$$(52) \quad \nabla \log \det(\nabla h(\gamma)) = \left(\frac{f''(|\gamma|)}{f'(|\gamma|)} + (k-1) \left[\frac{f'(|\gamma|)}{f(|\gamma|)} - \frac{1}{|\gamma|} \right] \right) \frac{\gamma^T}{|\gamma|}.$$

Plugging in the definitions of f , f' and f'' from (15), (37) and (38) for large x , we see that $f''(x)/f'(x)$ and $f'(x)/f(x)$ go to zero as x goes to infinity, and hence (52) goes to zero as $|\gamma|$ goes to infinity. Hence the numerator in (51) goes to zero. By (48) the denominator in (51) goes to infinity, and hence (51) holds. \square

LEMMA 6. *Let π_β be an exponentially light density on \mathbb{R}^k , and let h be defined by (5) and (15). Then π_γ defined by (4) has the property that*

$$(53) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \pi_\gamma(\gamma)}{|\nabla \pi_\gamma(\gamma)|}$$

(which is the limit superior in the curvature condition) is bounded above by

$$(54) \quad \limsup_{|\gamma| \rightarrow \infty} f'(|\gamma|) \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \log \pi_\beta(h(\gamma))}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|},$$

where the dots in both equations denote inner products.

PROOF. We always assume that π_β and π_γ are positive (Section 2.1), so we may take logs, obtaining

$$\frac{\nabla \log \pi_\gamma(\gamma)}{|\nabla \log \pi_\gamma(\gamma)|} = \frac{\nabla \pi_\gamma(\gamma)}{|\nabla \pi_\gamma(\gamma)|}.$$

Thus (53) can be rewritten as

$$\limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \log \pi_\gamma(\gamma)}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|} \frac{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|}{|\nabla \log \pi_\gamma(\gamma)|},$$

and then we can use Lemma 5 as

$$\limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \log \pi_\gamma(\gamma)}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|}.$$

If we expand $\nabla \log \pi_\gamma(\gamma)$ using (7), this is bounded above by the sum of

$$(55) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|}$$

and

$$(56) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{\gamma}{|\gamma|} \cdot \frac{\nabla \log \det(\nabla h(\gamma))}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|}.$$

It follows from Lemmas 2 and 5 that (56) is zero. Hence the limsup in (53) is bounded above by (55), which is equal to (54) since $\nabla h(\gamma)$ is symmetric and $\nabla h(\gamma)\gamma = f'(|\gamma|)\gamma$. \square

LEMMA 7. Let $a(\gamma)$ and $b(\gamma)$ be functions such that both a and b are positive and bounded away from zero and infinity as $|\gamma|$ goes to infinity. Then for f from (15), the fraction

$$(57) \quad f'(|\gamma|)^2 / \left(\frac{f(|\gamma|)^2}{|\gamma|^2} a(\gamma) + \left[f'(|\gamma|)^2 - \frac{f(|\gamma|)^2}{|\gamma|^2} \right] b(\gamma) \right)$$

is positive and bounded away from zero and infinity as $|\gamma|$ goes to infinity.

PROOF. The reciprocal of (57) is

$$\frac{f(|\gamma|)^2}{f'(|\gamma|)^2 |\gamma|^2} a(\gamma) + \left[1 - \frac{f(|\gamma|)^2}{f'(|\gamma|)^2 |\gamma|^2} \right] b(\gamma).$$

Since $a(\gamma)$ and $b(\gamma)$ are both positive and bounded away from zero and infinity for large $|\gamma|$, it is sufficient to show that

$$(58) \quad \frac{f(x)^2}{f'(x)^2 x^2}$$

is bounded away from zero and one for large x . For large x , it follows from (15) and (37) that (58) is equal to

$$\frac{[x + (x - R)^p]^2}{[1 + p(x - R)^{p-1}]^2 x^2},$$

which converges to $1/p^2$ as $x \rightarrow \infty$. Since we assume $p > 2$, we are done. \square

PROOF OF THEOREM 4. First, assume that condition (i) holds. By Lemma 6, it is enough to show that (54) is less than zero, and (54) is equal to, using (9),

$$(59) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{|\nabla \log \pi_\beta(h(\gamma))| f'(|\gamma|)}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|} \frac{h(\gamma)}{|h(\gamma)|} \cdot \frac{\nabla \log \pi_\beta(h(\gamma))}{|\nabla \log \pi_\beta(h(\gamma))|}.$$

Since π_β satisfies condition (3), there is an $\varepsilon > 0$ such that (59) is bounded above by

$$(60) \quad \limsup_{|\gamma| \rightarrow \infty} \frac{|\nabla \log \pi_\beta(h(\gamma))| f'(|\gamma|)}{|\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)|} (-\varepsilon).$$

Because $f'(|\gamma|)$ is strictly positive, the fraction in (60) is strictly positive for large $|\gamma|$, hence showing that this fraction's square is bounded away from zero is enough to show that (60) is less than zero, and condition (3) holds. Let

$$a(\gamma) = \frac{|\nabla \log \pi_\beta(h(\gamma))|^2}{|\nabla \log \pi_\beta(h(\gamma))|^2} = 1$$

and

$$b(\gamma) = \left(\frac{\nabla \log \pi_\beta(h(\gamma)) \cdot h(\gamma)}{|\nabla \log \pi_\beta(h(\gamma))| |h(\gamma)|} \right)^2.$$

Then, using (41) as in deriving (50), the square of the fraction in (60) is equal to (57). The Cauchy–Schwarz inequality bounds $b(\gamma)$ above by one, and condition (3) bounds $b(\gamma)$ away from zero. So by Lemma 7 the square of the fraction in (60) is positive and bounded away from zero as $|\gamma|$ goes to infinity. Because this fraction itself is positive, it must also be bounded away from zero as $|\gamma|$ goes to infinity. Hence the \limsup in (60) is negative and condition (3) holds for π_γ .

Now assume that condition (ii) holds and π_β is exponentially light, that is, there exist a $\beta_0 > 0$, $\varepsilon > 0$ and $M_1 > M_2 > 0$ such that for $|\beta| > \beta_0$,

$$\frac{\beta}{|\beta|} \cdot \nabla \log \pi_\beta(\beta) < -\varepsilon$$

and

$$M_2 < |\nabla \log \pi_\beta(\beta)| < M_1.$$

It follows that $1/|\nabla \log \pi_\beta(\beta)| > 1/M_1$ so π_β satisfies condition (i). \square

PROOF OF THEOREM 5. By (7) and the triangle inequality, $|\nabla \log \pi_\gamma(\gamma)|$ is bounded above by the sum

$$(61) \quad |\nabla \log \pi_\beta(h(\gamma)) \nabla h(\gamma)| + |\nabla \log \det(h(\gamma))|.$$

Hence it is sufficient to show that both of these terms are bounded as $|\gamma|$ goes to infinity.

It follows from (52) that the right-hand term in (61) is equal to

$$(62) \quad \frac{f''(|\gamma|)}{f'(|\gamma|)} + (k-1) \frac{f'(|\gamma|)}{f(|\gamma|)} - (k-1) \frac{1}{|\gamma|}.$$

For large y ,

$$(63) \quad f(y) = e^{by} - \frac{e}{3},$$

$$(64) \quad f'(y) = be^{by},$$

$$(65) \quad f''(y) = b^2 e^{by}.$$

So (62) is equal to

$$b + b(k-1) \frac{e^{b|\gamma|}}{e^{b|\gamma|} - e/3} - (k-1) \frac{1}{|\gamma|},$$

which clearly converges to bk as $|\gamma|$ goes to infinity, so the right-hand term in (61) is bounded for large $|\gamma|$.

It follows from (41) as in deriving (50) and from (9) that the square of the left-hand term in (61) is equal to the sum of

$$(66) \quad \frac{f(|\gamma|)^2}{|\gamma|^2} |\nabla \log \pi_\beta(h(\gamma))|^2$$

and

$$(67) \quad f'(|\gamma|)^2 \left[1 - \frac{f(|\gamma|)^2}{|\gamma|^2 f'(|\gamma|)^2} \right] \left(\frac{h(\gamma) \cdot \nabla \log \pi_\beta(h(\gamma))}{|h(\gamma)|} \right)^2.$$

It follows from (63) and (64) that the term in square brackets of (67) is positive and less than one for large $|\gamma|$. Since the other two terms in (67) are squares, (67) is nonnegative for large $|\gamma|$. Thus, applying the Cauchy–Schwarz inequality to the term in parentheses in (67), one bounds (67) above by

$$(68) \quad f'(|\gamma|)^2 |\nabla \log \pi_\beta(h(\gamma))|^2.$$

By $f(|\gamma|) = |h(\gamma)|$ and by (19), for $|\gamma|$ large (68) is bounded above by

$$\alpha^2 \frac{f'(|\gamma|)^2}{f(|\gamma|)^2},$$

which converges to $\alpha^2 b^2$ as $|\gamma|$ goes to infinity, and that finishes the proof that (61) is bounded for large $|\gamma|$ and the proof of the theorem. \square

REFERENCES

- BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. [MR0489333](#)
- BROOKS, S., GELMAN, A., JONES, G. L. and MENG, X.-L., eds. (2011). *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. [MR2742422](#)
- CHAN, K. S. (1993). On the central limit theorem for an ergodic Markov chain. *Stochastic Process. Appl.* **47** 113–117. [MR1232855](#)
- CHAN, K. S. and GEYER, C. J. (1994). Comment on “Markov chains for exploring posterior distributions.” *Ann. Statist.* **22** 1747–1758.
- DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. [MR0520238](#)
- FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. [MR2604704](#)
- GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. [MR1141740](#)
- GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.
- GEYER, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7** 473–511.
- GEYER, C. J. (2011). Introduction to MCMC. In *Handbook of Markov Chain Monte Carlo* (S. P. Brooks, A. E. Gelman, G. L. Jones and X. L. Meng, eds.). Chapman & Hall/CRC, Boca Raton.
- GEYER, C. J. and JOHNSON, L. T. (2012). mcmc: Markov Chain Monte Carlo. R package version 0.8. Available at <http://CRAN.R-project.org/package=mcmc>.
- GEYER, C. J. and MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.* **21** 359–373. [MR1310082](#)
- GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics*. Chapman & Hall, London. [MR1397966](#)

- GORDIN, M. I. and LIFŠIC, B. A. (1978). Central limit theorem for stationary Markov processes. *Dokl. Akad. Nauk SSSR* **239** 766–767. [MR0501277](#)
- GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. [MR1380810](#)
- HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
- HOBERT, J. P. and GEYER, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J. Multivariate Anal.* **67** 414–430. [MR1659196](#)
- JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. [MR1731030](#)
- JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. [MR1890063](#)
- JARNER, S. F. and ROBERTS, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. [MR2396939](#)
- JARNER, S. F. and TWEEDIE, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9** 559–578. [MR1996270](#)
- JOHNSON, A. A. and JONES, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Stat.* **4** 313–333. [MR2645487](#)
- JONES, G. L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320. [MR2068475](#)
- JONES, G. L. and HOBERT, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann. Statist.* **32** 784–817. [MR2060178](#)
- KIPNIS, C. and VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. [MR0834478](#)
- ŁATUSZYŃSKI, K., MIAOJEDOW, B. and NIEMIRO, W. (2012). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli*. To appear.
- ŁATUSZYŃSKI, K. and NIEMIRO, W. (2011). Rigorous confidence bounds for MCMC under a geometric drift condition. *J. Complexity* **27** 23–38. [MR2745298](#)
- MAIGRET, N. (1978). Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **14** 425–440. [MR0523221](#)
- MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. [MR1389882](#)
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **31** 1087–1092.
- MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2509253](#)
- NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge Tracts in Mathematics **83**. Cambridge Univ. Press, Cambridge. [MR0776608](#)
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Statist. Sci.* **22** 59–73. [MR2408661](#)
- PAPASPILIOPOULOS, O. and ROBERTS, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Ann. Statist.* **36** 95–117. [MR2387965](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25 (electronic). [MR1448322](#)
- ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. [MR2095565](#)
- ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. [MR1440584](#)

- ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. [MR1399158](#)
- ROCKAFELLAR, R. T. and WETS, R. J. B. (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Springer, Berlin. [MR1491362](#)
- ROSENTHAL, J. S. (1995a). Analysis of the Gibbs sampler for a model related to James–Stein estimators. *Stat. Comput.* **6** 269–275.
- ROSENTHAL, J. S. (1995b). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. [MR1340509](#)
- ROY, V. and HOBERT, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 607–623. [MR2370071](#)
- STROMBERG, K. R. (1981). *Introduction to Classical Real Analysis*. Wadsworth International, Belmont, CA. [MR0604364](#)
- TAN, A. and HOBERT, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. [MR2598033](#)
- TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. [MR0898357](#)
- TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. [MR1329166](#)

GOOGLE INC.
1600 AMPHITHEATRE PARKWAY
MOUNTAIN VIEW, CALIFORNIA 94043
USA
E-MAIL: ltjohnson@google.com

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
313 FORD HALL
224 CHURCH ST. SE
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: charlie@stat.umn.edu